Supplementary Materials for

Parallel molecular evolution in an herbivore community

Ying Zhen, Matthew L. Aardema, Edgar M. Medina, Molly Schumer and Peter Andolfatto*

correspondence to: pandolfa@princeton.edu

This PDF file includes:

Table of Contents
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S11
Supplementary References (36-80)

Other Supplementary Materials for this manuscript includes the following:

Databases S1: mRNA-seq data for 21 species surveyed in our study.
http://genomics-pubs.princeton.edu/insect_genomics/data.shtml

Table of Contents

Materials and Methods

Survey of ATPα in 21 insect taxa. 4-5

Evolutionary analyses. 5

Protein structure and molecular docking simulations. 5-6

Gene Expression Assays. 6

Supplementary Text

1. Supplementary Methods.

    1.1. Proof of concept for mRNA-seq as a gene discovery tool; Example: RNAseq data for *D. melanogaster* and *D. plexippus* 6-7

    1.2 Details of *de novo* assembly for each species 7-11

    1.3 Establishing duplicates versus alleles in the dogbane beetle, milkweed stem weevil and milkweed bugs 11-12

2. Estimating branch lengths and the ages of ATPα1 duplicates.

    2.1. Estimation of branch lengths 12

    2.2. Age of ATPα1 duplicates in beetles 12-13

    2.3. Age of ATPα1 duplicates in true bugs 13

3. Summary of molecular docking simulations. 13-14

4. Copy-specific ATPα1 qPCR primers and their efficiencies. 15

5. Tissue-specific expression of ATPα1 copies A, B and C in large and small milkweed bugs. 15

Supplementary Figures

Fig. S1. Summary of sites in ATPα involved in ouabain-binding based on site-directed mutagenesis studies and protein structure data. 16

Fig. S2. Phylogenetic relationships of ATPα orthologs and paralogs. 17

Fig. S3. Visual summary of molecular docking simulations for each of the substitutions in Table S2. 18-24

Fig. S4. Tissue-specific relative expression of ATPα1 duplicates in the large milkweed bug and the small milkweed bug. 25

Fig. S5. Patterns of substitution at functional sites identified in mutagenesis and structural studies extended to include ATPα2. 26

Fig. S6. Manual reconstruction of *C. auratus* ATPα1 copies. 27

Fig. S7. Establishing duplicates in *C. auratus*. 28

Fig. S8. Establishing duplicates in *O. fasciatus.* 29

Supplementary Tables

Table S1. List of species and collection information. 30

Table S2. Summary of effects of each substitution associated with use of Apocynaceae on docking of ouabain onto the pig ortholog of ATPα. 31

Table S3. Summary of dockings of ouabain onto predicted structures of native ATPα1 for eleven native proteins with ≥1 "large-effect" substitutions and eight with no such substitutions. 32

Table S4. Summary of dockings of ouabain onto "revertant" versions of native ATPα1. 33

Table S5. Functional substitutions observed in duplicates of ATPα1 of large and small milkweed bugs**.** 34

Table S6. Taxa in this study with references for phylogenetic relationships. 35

Table S7. Details of mRNA-seq data collected. 36

Table S8. List of reference genome sequences and accession numbers used in this study. 37

Table S9. Summary of copy-specific ATPα1 qPCR primers and their qPCR efficiencies. 38

Table S10. Proportion of total branch length sampled for taxa "using" and "not using" Apocynaceae host plants. 39

Table S11. Of lineages associated with Apocynaceae host plants, the proportion of total branch length sampled for lineages with "duplicated" and "not duplicated" versions of ATPα1. 40

Supplementary References 41-42

Materials and Methods

Survey of ATPα in 26 insect taxa

Eleven Apocynaceae-feeding insects were collected on *Asclepias syricia* and *Apocynum cannabinum* in Princeton, NJ and stored in RNAlater (Qiagen, Valencia, CA) at -80 °C. Three additional species (*Danaus gilippus*, *Danaus eresimus*, *Lycorea halia*) that use Apocynaceae hosts to varying extents were a kind gift from A. Briscoe (Table S1). Seven non-Apocynaceae feeding outgroup species were also collected (Table S1). Total RNA was extracted using Tri reagent (Ambion, Grand Island, NY) according to the manufacturer's instructions. The RNA was purified using RNeasy Mini kit (Qiagen), and was treated with RNase-free DNase (Qiagen) to remove DNA contamination.

We surveyed the near complete ATPα proteins in these species in order to fully integrate comparative sequence data with site-directed mutagenesis (SDM) and crystallographic structure data. A traditional approach to surveying a gene of interest in non-model organisms is PCR using degenerate primers (e.g. *20, 22*). However, this approach is not only labor-intensive, particularly for a large protein like ATPα, but could be biased by information used for primer design, particularly when multiple copies of a gene are present in the genome (*36*). We instead opted to use a less biased, yet still cost-effective, "gene discovery" approach based on mRNA-seq (*37*). Combined with multiplexing, this approach allowed us to survey the ATPα protein coding sequences efficiently in multiple taxa. After confirming the utility of this approach using previously published mRNA-seq data from *Drosophila melanogaster* (Supplementary text 1.1), we assembled near full-length transcripts of ATPα from 14 Apocynaceae-feeding taxa and 12 outgroups (the availability of full genome sequences provided information for three more outgroups for a total of 15, Figure 1, Supplementary text 1.2).

Libraries were prepared using Illuminia's mRNA-seq Sample Prep Kit (Illumina, San Diego, CA) following the manufacturer's protocol with slight modifications. Specifically, we prepared each library with custom barcoded linkers that allowed us to multiplex them as needed (1-16 plex). Libraries were sequenced in several lanes on an lllumina Genome Analyzer II and a HiSeq 2000 at the Lewis-Sigler Institute Microarray Facility (Princeton University, Princeton, NJ). In retrospect, with advances in sequencing technology during the course of this project, the entire experiment (21 taxa) could have been completed in a single paired-end HiSeq run (i.e. ~10 million pairs per taxa). Reads were parsed by barcode and the numbers of reads for each species are listed in Table S7. Raw reads were trimmed to Phred quality score ≥20 prior to *de novo* assembly using Velvet/Oases (http://www.ebi.ac.uk/~zerbino/oases/).

In addition to the 21 taxa described above, we also generated our own *de novo* assemblies of ATPα1 for *D. melanogaster* (as proof of concept) and 5 more outgroup species using previously published mRNA-seq data (Table S8). The assembly for each of the species is detailed in Supplementary text 1.2. Briefly, the assembly was straightforward for 17 of the 21 species in which ATPα1 appeared to be present in just one copy. In each case, we used the predicted protein sequences for the ATPα1 of *D. melanogaster* to query the assembled transcripts using BLAST (`tblastn`). Of the transcripts identified, the longest was typically chosen. Multiple transcripts in these cases generally reflected what appear to be minor splice-variants of the protein. For the

remaining four species, duplicate copies were detected and the assembly required manual intervention to ensure proper assembly (see Supplementary text 1.2 and Fig. S6). In these cases, raw reads were assembled to Velvet contigs to manually extend their ends and we were able to link contigs when there was sufficient confidence (supported by ≥3 raw reads). Finally, we designed ortholog-specific PCR primers to fill in the remaining gaps and confirm the phase of our final assembly.

All raw sequence data produced for this project has been posted on a publicly-accessible server (http://genomics-pubs.princeton.edu/insect_genomics/data.shtml). Of the 21 taxa surveyed, all are non-models and 20 (i.e. all except monarch) have no prior genome scale data. Our goal of assembling of ATPα from each taxa was modest and the data could be mined to ask a variety of additional questions in these non-model species. The assembled sequences of ATPα used in our analysis have been submitted to Genbank under Accession numbers JQ771496-527.

Evolutionary and statistical analyses

In addition to the 26 sequences generated above, we also used previously published genome reference sequences for three additional species (see Table S8). Alignments of the ATPα cDNA and protein sequences were performed using Geneious Pro 5.3.4 software with manual adjustment. Due to their poor alignment, the first 20 amino acids (relative to the sheep reference sequence, Table S8) were removed prior to analysis. We used previous studies to establish phylogenetic relationships among the surveyed taxa (Figure 1, Table S6). In particular, we focused on patterns of substitution at a subset of sites (Fig. S1) that were previously identified as either having a direct functional role in the affinity of ATPα for the cardenolide ouabain (by SDM studies) or were identified in protein structure studies as being likely to interact with ouabain. For each site we used standard parsimony criteria to establish the state of ancestral nodes. Branch lengths were estimated using PAML (38; Supplementary text 2.1). For statistical analysis of clustering of mutations, we formulated a two-tailed binomial test to evaluate the probability that X or more of 33 substitutions associated with feeding on Apocynaceae occur at 1 of 35 potential targets; X=12 for site Q111 and X=6 for site N122. To evaluate whether substitutions are clustered in the H1-H2 extracellular domain, we calculate the probability that 24 or more of the 33 substitutions occur at 7 of 35 potential targets.

Protein structure and molecular docking simulations

We used molecular docking simulations to examine the effects of particular amino acid substitutions on the affinity of ATPα for the cardenolide ouabain. Specifically, we used AutoDock Vina v.1.1.2 (39) for docking calculations against the crystal structure of ATPα (ATP1A1) of *Sus scrofa* (the domestic pig) bound to ouabain in the E2P form (PDB Accession Number 3N23; 31). The pig structure is used because it is the only available structure of Na^+,K^+-ATPase bound to ouabain in the high affinity (i.e. E2P) state. To validate docking analyses, we extracted ouabain from the co-crystal structure and performed re-docking of the ouabain against 3N23A using the induced fit shape of the receptor and flexible side-chains based on the β-factors of the crystal structure. The best results (defined as the docking closest to the coordinates of ouabain in the co-crystal structure) were obtained with the induced fit shape of the receptor. To determine the effects of each amino acid substitution, we simulated mutations in ATPα followed by

receptor preparation and docking as performed in the re-docking analysis. The results of these analyses are detailed in Supplementary text 3; Fig. S3.

For all docking analyses, the structure of ouabain was downloaded from PubChem Substance and Compound database (CID 439501) and prepared for docking with 200 steps of rigid-body minimization in PyRx v.0.8 (http://pyrx.sourceforge.net/). The receptor preparation for docking was carried out for the structure in absence of ouabain by 100 steps of rigid-body minimization and with Dock Prep using default parameters as implemented in UCSF Chimera v.1.5.3 (40). Gasteiger charges were computed using ANTECHAMBER (41). The pdbqt files for AutoDock Vina were prepared using AutoDock Tools (ADT) as implemented in MGLTools 1.5.4 (http://autodock.scripps.edu/resources/adt). The co-crystal coordinates of ouabain in 3N23 were used to locate a grid box of 20x20x30 Angstroms for the docking searching space and an exhaustiveness of 10 was used in AutoDock Vina. The results were visualized in and prepared for figures in UCSF Chimera.

Gene Expression Assays

For steady-state mRNA expression analysis of ATPα1 duplicates, we designed copy-specific qPCR primers (Supplementary text 4, Table S9). Three to five individuals from each species were assayed separately. For each individual, we sampled three tissues for RNA extraction: whole head, digestive tract (from the esophagus to rectum with Malpighian tubules and gut contents removed), and thoracic muscle. Dissected tissues were immediately placed in RNAlater (Ambion). Total RNA was extracted using a standard Trizol protocol (Invitrogen, Grand Island, NY), treated with RNase-free DNase, and reverse-transcribed into first-strand cDNA using the GoScript Reverse Transcription System (Promega, Madison, WI) according to manufacturer's protocol. qPCR was done using SYBR Green PCR Master Mix (Applied Biosystems®, Grand Island, NY) on StepOnePlus Real-Time PCR Systems (Applied Biosystems®). Differences in the Ct value between different copies of ATPα1 were calculated for each sample and variation among tissues was evaluated by ANOVA followed by a Tukey-Kramer test (implemented in the R statistical package, http://www.r-project.org).

Supplementary Text

1. Supplementary Methods

1.1. Proof of concept for mRNA-seq as a gene discovery tool; Example: RNAseq data for *D. melanogaster* and *D. plexippus*.

We created a de novo assembly using 4.4 million, 100 bp paired-end Illumina reads from a library prepared from ovaries of *D. melanogaster* strain Oregon-R (SRR384962, http://www.ncbi.nlm.nih.gov/sra/SRX109311), using Velvet/Oases with kmer length set to 31 (42; http://www.ebi.ac.uk/~zerbino/oases/). Based on its similarity to the predicted ATPα1 protein of *Bombyx mori* (BGIBMGA005058-TA), we recovered one copy of ATPα1 in four minimally overlapping transcript fragments. The predicted protein is 92.5% identical to the predicted ATPα1 of *B. mori* and 99.3% (all but 1 substitution are possibly due to alternative splicing variants) identical to ATPα1 of the *D. melanogaster* reference genome sequence (AF044974).

We also assembled 28.3 million 101 bp single-end reads from a monarch butterfly (*Danaus plexippus*, see Table S7) with Velvet/Oases using kmer length set to 31. Using the *D. melanogaster* ATPα1 protein sequence (AF044974) to query the assembled transcripts, we recovered 8 transcripts, 3 of which had full length CDS and appear to be minor splice variants. We chose the longest of these (4935 bp) transcripts that included the full length CDS. The predicted protein for Monarchs is 92.6% and 97.7% identical to the predicted proteins for *D. melanogaster* and *B. mori*, respectively. Our predicted protein is also almost identical to DPGLEAN09831 of the Monarch reference genome (with the exception of two small regions that likely represent alternative splicing forms), but only ~50% identical to two other copies in the Monarch genome (DPGLEAN02030 and DPGLEAN02739, see Fig. S2). The latter copies belong to a distinct clade of ATPα (ATPα2), which we discuss below. These did not appear in our assembly, likely because they have low levels of expression (32).

1.2 Details of *de novo* assembly for each species

All reads (75-125 bp) were quality trimmed to Phred QV ≥ 20 and C ≥ 20 (the number of contiguous bases) using the script TQSfastq.py (http://code.google.com/p/ngopt/source/browse/trunk/SSPACE/tools/TQSfastq.py). For all species, we used either Velvet (42) or the combination of Velvet and Oases (http://www.ebi.ac.uk/~zerbino/oases/) to produce a de novo assembly of mRNA-seq data described in Table S7. In each case, we identified ATPα orthologs using BLAST (tblastn, http://blast.ncbi.nlm.nih.gov/) and the *D. melanogaster* ATPα1 protein sequence as the reference query. In many of our assemblies, we identified a second highly diverged locus belonging to the ATPα2 clade, an ancient duplication of ATPα in insects (see Fig. S2). When we assembled multiple transcripts for a given locus with Oases, the longest with complete CDS was used. Post assembly, all reads were assembled back to the identified ATPα1 ortholog and the CDS region was checked manually for error and polymorphisms using Geneious (http://www.geneious.com/). If more than one of these transcripts was identified as an alternative splicing variant, the one with the highest number of uniquely mapped reads to the alternatively spliced region was used. In some cases (discussed below), Velvet/Oases transcripts were very short in length, or were inconsistent with the patterns of reads mapped. In these cases, we reverted to a more "manually intensive" approach based on Velvet assemblies (described in detail below). Specific details of the assembly for each species are (in the order that they appear in Fig. 1) as follows:

Danaus gilippus: We assembled 1.7 million, 100-101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One locus with two transcripts was identified that included the full length CDS, likely reflecting alternative-splicing variants near the 3' end of the protein.

Danaus eresimus: We assembled 5.9 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. Three transcripts were found, two of which minimally overlapped and a third that was separated from the other two by a small gap (~23 bp). This gap was filled using PCR with primers designed from flanking regions (Forward: ACGCTTCGTGCTGAGGATCTGG; Reverse: ACCGTGTTGTCACCGCAGCA). Total RNA from *D. eresimus* was reverse transcribed to cDNA using M-MLV reverse transcriptase (Promega, Madison, WI) following

manufacturer's instructions. We used PrimeSTAR HS DNA polymerase (Takara, Mountain View, CA) and cDNA as template for PCR. The product was cleaned using the QIAquick PCR Purification Kit (Qiagen), 3'A-tailed using Apex Taq polymerase (Genesee scientific, San Diego, CA), and cloned into TOPO TA vector (Invitrogen), following the manufacturer's instructions. One of the clones was picked and sequenced using M13F primer. The resulting sequence was assembled together with the previous transcripts identified to produce a near full length consensus CDS.

Lycorea halia atergatis: We assembled 16.1 million 98 bp single-end and 1.5 million 100-101bp paired-end reads using Velvet/Oases with kmer length set to 31. One locus with three equal-length transcripts and full length CDS were identified. These likely represent alternative splicing variants near the 3' end of the protein.

Limenitis archippus: We assembled 3.9 million 100 bp single-end reads using Velvet/Oases with kmer length set to 31. One locus with a single transcript including full length CDS was identified.

Heliconius melpomene: We assembled previously published 11.6 million 125 bp single-end reads (only read 1 of paired-end data were used) using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Papilio glaucus: We assembled 25.9 million 101 bp single-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Euchaetes egle: We assembled 5.9 million 95-100 bp single-end and 7 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Cycnia tenera: We assembled 5.6 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One locus including the full length CDS was identified, which had two alternative splicing variants.

Lophocampa caryae: We assembled 5 million 100 bp single-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Trichordestra legitima: We assembled 4.7 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Chrysochus auratus: We used 3.3 million 95 bp single-end and 3.9 million 100-101 bp paired-end reads. We first attempted an assembly using Velvet/Oases with kmer length set to 31. The locus with the highest similarity to ATPα1 contained 7 transcripts with full length CDS. Upon mapping the raw reads back to these transcripts, it became apparent that multiple copies of ATPα1 may be present, and that these did not appear to belong to the ATPα2 family (Fig. S2). We thus decided to revert to Velvet assemblies with kmer lengths set to 21 and 31 and exp_cov set to auto. We merged Velvet contigs from the multiple assemblies. Fourteen distinct Velvet contigs showed high levels of similarity to *D. melanogaster* ATPα1, with strong evidence for distinct two copies of ATPα1 (i.e. two distinct alleles at most of the sites, Fig. S6A). With these contigs, we performed an iterative manual contig extension using the raw reads to span gaps and increase the extent of overlap of contigs to unambiguously establish the phase of the duplicates (Fig. S6B and C). We established two near full-length ATPα1 orthologs

(named ATPα1A and ATPα1B, respectively). A description of how we distinguish duplications from alleles of a given duplicate is given below.

Labidomera clivicollis: We assembled 6.9 million 95-100 bp single-end reads using Velvet /Oases with kmer length set to 31. One transcript with near full length CDS was identified as orthologous to ATPα1.

Plagiodera versicolora: We assembled 1.9 million 100 bp single-end reads using Velvet/Oases with kmer length set to 31. Two non-overlapping transcripts were identified with a 24 bp gap between them. This 24 bp gap was first filled with *D. melanogaster* ATPα1 to create a mosaic reference sequence. By assembling all the trimmed raw reads to this mosaic reference, we found three raw reads that spanned this gap to produce a full length ATPα1 ortholog.

Tetraopes tetraophthalmus: We assembled 8.3 million 95-100 bp single end reads, using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1. A second highly diverged locus with partial CDS had closest homology to the ATPα2 clade.

Megacyllene robiniae: We assembled 4.2 million 100 bp single-end reads using Velvet /Oases with kmer length set to 31. One locus with full length CDS was identified as orthologous to ATPα1, with two equal-length alternative splicing transcripts. A second highly diverged locus with partial CDS was also identified that had closest homology to the ATPα2 clade.

Rhyssomatus lineaticollis: We used 3.4 million 95 bp single-end and 4.3 million 100-101 bp paired-end reads. Similar to *C. auratus*, our first attempt at an assembly using Velvet/Oases (kmer length = 31) resulted in a large number of transcripts. Upon mapping the raw reads back to these transcripts, it became apparent that multiple copies of ATPα1 may be present. Thus, in this case, we also reverted to using Velvet assemblies with kmer length set to 21, 27, 31 or 35 and exp_cov set to auto. Fifteen distinct Velvet contigs showing high sequence identity with *D. melanogaster* ATPα1 suggested the presence two distinct copies of ATPα1. With these contigs, we performed an iterative manual contig extension using the raw reads to span gaps and increase the extent of overlap of contigs to unambiguously establish the phase of the duplicates (as we did with *C. auratus*). We recovered two near full-length ATPα1 orthologs (named ATPα1A and ATPα1B, respectively).

Cyrtepistomus castaneus: We assembled 12.4 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1. A second highly diverged locus with full length CDS was also identified that had closest homology the ATPα2 clade (as discussed above).

Pogonus chalceus: We assembled previously published 25.9 million 101 bp single-end reads (only first 10 million reads in read 1 of paired-end data were used) using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Oncopeltus fasciatus: We assembled 7.9 million 95-100 bp single-end and 27.1 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. Two transcripts were identified as two distinct full length and near full length copies of ATPα1 (named ATPα1A and ATPα1B, respectively). However, two additional short transcripts assembled to the 3' end of ATPα1 suggested the existence of two more ATPα1 copies. To improve the assembly of these contigs, we used Velvet contigs (constructed with kmer

length set to 21, 27 or 31, and exp_cov set to auto). We merged the resulting contigs as we did for *Chyrosochus* and *Rhyssomatus*. Fourteen distinct Velvet contigs were identified with a high degree of similarity to ATPα1. From these, contigs belonging to ATPα1A and ATPα1B were identified and removed. From the remaining contigs, we designed PCR primers from contigs mapping to the 5' and 3' ends of the protein. Sequencing these products allowed us to bridge together remaining contigs, producing a third, near full length, copy (ATPα1C). Three short contigs remained that did not belong to any of the orthologs we identified, which suggested that more copies might exist. In addition, a second highly diverged locus with partial CDS was also identified and had closest homology to the ATPα2 clade.

Lygaeus kalmii: We assembled 9.5 million 95-100 bp single-end and 21.2 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as an ATPα1 ortholog (ATPα1B, which is orthologous to ATPα1B of *Oncopeltus*). However, three additional short transcripts assembled to ATPα1 suggested the existence of more ATPα1 copies. To improve the assembly we used Velvet contigs (constructed with kmer length set to 21, 27 or 31, and exp_cov set to auto). We identified 19 distinct Velvet contigs with high similarity to ATPα1. Of these we removed contigs identified as belonging to ATPα1B. With the remaining contigs, we performed an iterative manual contig extension using the raw reads to span gaps and increase the extent of overlap of contigs to unambiguously establish the phase of the duplicates (as we did with *C. auratus*). This allowed us to recover a near full length copy of ATPα1A (orthologous to ATPα1A of *Oncopeltus*). From the remaining contigs, we designed PCR primers from contigs mapping to the 5' and 3' ends of the protein. Sequencing these products allowed us to bridge together remaining contigs, producing a third, near full length, copy (ATPα1C). Of the remaining unassembled Velvet contigs, two short contigs suggested the possible existence of additional ATPα1 copies. In addition, we identified two Velvet contigs that included near full length CDS belonging to the ATPα2 clade. We could link these together with support from two pairs of paired-end reads (though the intervening 46 bp gap between them had only one read coverage).

Boisea trivittata: We assembled 4.9 million 100 bp single-end reads using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Cimex lectularius: We assembled previously published 10 million 74 bp single-end reads (only first 10 million reads in read 1 of paired-end data were used) using Velvet/Oases with kmer length set to 31. One transcript with full length CDS was identified as orthologous to ATPα1.

Aphis nerii: We assembled 4.9 million 101 bp paired-end reads using Velvet/Oases with kmer length set to 31. *Acyrthosiphon pisum* ATPα1 (XM_001948888) was used as a reference. One locus with full length CDS was identified as orthologous to ATPα1. The longest transcript had an insertion near 5' end and a much shorter CDS comparing to the reference (possibly due to alternative splicing). We used the second-longest transcript at this locus for a more comparable CDS. A second highly diverged locus with partial CDS was also identified and had closest homology to the ATPα2 clade (as discussed above).

Bemisia tabaci: We assembled previously published 21.9 million 75 bp paired-end reads using Velvet/Oases with kmer length set to 21. Three transcripts fragments were

identified, two of which minimally overlapped and a third that was separated from the other two by a small gap (~20 bp). The gap was filled by iterative manual contig extension of the end of one transcript to give the near-full length consensus CDS. Three additional short transcripts assembled to ATPα1 suggested the existence of more ATPα1 copies.

Diaphorina citri: We assembled previously published 100 bp paired-end reads using Velvet/Oases with kmer length set to 31. One locus with full length CDS was identified as orthologous to ATPα1. Another two highly diverged locus with full length or near full length CDS were also identified that had closest homology the ATPα2 clade (as discussed above).

1.3. Establishing duplicates versus alleles in the dogbane beetle, milkweed stem weevil and milkweed bugs.

Chrysochus auratus: Copy-specific primers were designed near the ends of each duplicate to PCR confirm the correct assembly (C_auratus_ATPα1A, Forward: TCACGAAGGAAAGTTAAGAAGATTAGG, Reverse: TGTTCTAACCAACCACCTGGGC; C_auratus_ATPα1B, Forward: GTCACGAAGGAAAGTTAAGAAGGT, Reverse: TCGTTCTAACCAACCTCCTGGAT). cDNA synthesis and PCR was conducted as described above, and cleaned PCR products were Sanger sequenced using the PCR primers. Additionally, trimmed raw reads were assembled to the sequences of the two copies and manually checked for polymorphisms. The existence of more than two alleles at some sites within each duplicate indicated these two copies were indeed duplicated genes rather than two alleles in one locus (Fig. S7). In follow-up expression studies, we confirmed that each duplicate could be amplified separately with copy-specific primers (Table S9) in at least three separate individuals.

Rhyssomatus lineaticollis: The low level of divergence between these duplicates prevented us from designing copy-specific primers that would amplify the entire CDS. Thus, primers were designed near the ends of these orthologs to PCR both orthologs and confirm the correct assembly of these orthologs (Forwad: ATGACGAGCATGGCCGTTCTGA, Reverse: GCCACCAGGCGGCAGTTGAA). PCR products were cleaned, and cloned as described above. Several clones were picked and sequenced using the M13F and M13R primers to confirm the existence of four alleles. Two additional internal primers were used to sequence through the full length CDS (R_lineaticollis_ATPα1A, Forward: TGTTGTCATTAGTTGCGGAGAC, Reverse: TCGGATTCGGCCTCCTCA; R_lineaticollis_ATPα1B Forward: TGGTCATTAGTTGCGGCGAT, Reverse: TCGGATTCGGCCTCTTCG). We also assembled trimmed raw reads to these two copies and manually checked for polymorphisms. The existence of more than two alleles at some sites indicated these two copies were indeed duplicated genes rather than two alleles at one locus (similar to the results for *C. auratus* in Fig. S7). In follow-up expression studies, we confirmed that each duplicate could be amplified separately with copy-specific primers (Table S9) in at least three separate individuals.

Oncopeltus fasciatus: All three copies were further confirmed by copy-specific primers (O_fasciatus_ ATPα1A, Forward: TAACCCTGAGACGGGACTCT, Reverse:

CAGGATATCGACGCAGATAGTATT; O_fasciatus_ATPα1B, Forward: TAACCCTGAAACGGGACTTA; Reverse: AACCACCAGGATTTCTGCGA; O_fasciatus_ ATPα1C, Forward: TGGGCTGACCCATGCAAAAGCA, Reverse: AATGAGCCTTCCTTCTTCGACG). Cleaned PCR products for ATPα1A and ATPα1B were directly sequenced, and all were cloned as described above. Individual clones harboring a single copy were sequenced and confirmed matching of our assembly, and these clones are further used as a template to test qPCR primer specificity and efficiency (described below). Further, trimmed raw reads were assembled to the three copies and we identified copy-specific polymorphisms in each copy. This confirmed that these copies were indeed duplicated genes rather than alleles (see Fig. S8). In follow-up expression studies, we confirmed that each duplicate could be amplified separately with copy-specific qPCR primers (Table S9) in at least four separate individuals.

Lygaeus kalmii: All three copies were further confirmed by copy-specific primers near the ends of these duplicates (L_kalmii_ATPα1A, Forward: CTTCTTCTGACCGAGCGCAA, Reverse: ACAAGAGTTAAGCTCAAGAAGAAAGC; L_kalmii_ATPα1B, Forward: TGTGCTTCTACTGACAATCCGA, Reverse: ACCAGTTTAAAAGAGAAATTTTGCTGA; L_kalmii_ATPα1C, Forward: TTAGATGATCTGAAGCAGGAA, Reverse: GCTCTTTAGTAGTATGTTTCTTG). Cleaned PCR products for ATPα1A and ATPα1B were directly sequenced, and all were cloned as described above. Individual clones harboring a single copy were sequenced and confirmed matching of our assembly, and these clones are further used as a template to test qPCR primer specificity and efficiency (described below). Further, trimmed raw reads were assembled to the three copies and we identified copy-specific polymorphisms in each copy. This confirmed that these copies were indeed duplicated genes rather than alleles. This confirmed that these copies were indeed duplicated genes rather than alleles (similar to Fig. S8). In follow-up expression studies, we confirmed that each duplicate could be amplified separately with copy-specific qPCR primers (Table S9) in at least five separate individuals.

2. Estimating branch lengths and the ages of ATPα1 duplicates.

2.1. Estimation of branch lengths.

We separated the taxa into separate clades representing distinct insect orders and did a separate analysis for each. In each case, we used the PAML script codeml (*38*) to estimate dS (the per site rate of substitution at synonymous sites) along particular branches using a prior tree based on established cladistic relationships among taxa (see Table S6). Parameters were set as follows: CodonFreq = 2; clock = 0; model = 0; fix_kappa = 0; kappa = 1.6; fix_omega = 0; ncatG = 1. We further partitioned branches into "using" and "not using" Apocynaceae (Table S10). For lineages associated with the use of Apocynaceae, we further partitioned branch length into lineages in which ATPα1 is "duplicated" or "not duplicated" (Table S11). In each case the relative amounts of branch length in these two types of lineages were used as the expected proportions in binomial tests comparing the observed number of substitutions.

2.2. Age of ATPα1 duplicates in beetles.

We calibrated the rate of evolution based on a splitting time of 74 Mya for the subfamilies Chrysomelinae and Eumolpinae (43). dS between *C. auratus* ATPα1 A and B copies is 0.446, which is ~33% of the average dS (1.369) between Chrysomelinae (*P. versicolora* and *L. clivicollis*) and Eumolpinae (*C. auratus*). Thus, we infer the divergence time for the *C. auratus* duplicates is ~24 Mya (0.3258*74 Mya). We could not perform a similar calibration specifically for weevils (*R. lineaticollis*). However, if we assume that the rate of evolution is close to that for Chrysomlids, we infer that dS between A and B copies in *R. lineaticollis* (0.168) corresponds to ~9 Mya.

2.3. Age of ATPα1 duplicates in true bugs.

The only available calibration for dS in Hemiptera is for aphids which estimates that *A. nerii* and *A. pisum* diverged ~65 Mya (*43*). Since we estimate dS to be 0.25 between these species, this implies ~5-fold slower evolution in Aphids compared to Coleoptera. Despite the closer evolutionary relationship between the aphids and the lygaeid bugs, we used a calibration based on known beetle divergence times due to the many major differences in life-histories between aphids and the Lygaeidae (e.g. the number of generations per year, obligate sexuality, etc.). dS between *O. fasciatus* and *L. kalmii* is 0.518, 0.412 and 0.577 for ATPα1 copies A, B and C, respectively. In contrast, average dS between duplicates A and B is 1.11 while average dS between (A,B) and copy C is 2.30. These estimates strongly suggest that both duplications predated the divergence of *O. fasciatus* and *L. kalmii*. Using the rate of evolution inferred for Coleoptera, we estimate that *O. fasciatus* and *L. kalmii* split ~30 Mya, whereas the A and B duplication arose ~60 Mya and C split from A and B ~125 Mya.

3. Summary of molecular docking simulations.

To explore the likely effects of specific amino acid substitutions on the interaction between Na^+,K^+-ATPase and cardenolides, we carried out a series of molecular docking simulations using Autodock Vina (39; see Materials and Methods):

Approach 1: In the first set of simulations (Table S2), we asked about the effects of individual amino substitutions observed at functionally important sites in a broad range of Apocynaceae specialist taxa (Figure 1) on the docking position of the cardenolide ouabain. To do this, we started with the co-crystal structure of pig (*Sus scrofa*) ATP1A1 in the E2P form bound to the cardenolide ligand ouabain (PDB 3N23). We compared the best re-docking of ouabain onto the pig structure to the best docking of ouabain onto the pig structure modified to include one of 18 possible substitutions. We define the best docking as the docking among the top 10 highest affinity dockings that was closest to the co-crystal coordinates of ouabain in PDB 3N23 (measured as Root Mean Square Distance, RMSD, in Angstroms). As the resolution of PDB 3N23 is 4.6 Angstroms (Å), we considered a substitution causing a displacement of >4.6 Å as having a "large" effect. Of the 18 substitutions simulated (Table S2), we identified five substitutions of large effect (C104Y, N122H, N122H, R880S, and R972Q), and two of these (C104Y and N122H) have known effects on ouabain-sensitivity (Fig. S1).

Approach 2: ATPα1 is a highly conserved protein, particularly at sites implicated in ouabain-binding (Figure 1). There may be a concern that above structural studies for individual substitutions observed in our Apocynaceae-feeding insects are based on the

structural information and sequence for a vertebrate (pig) ortholog of the protein. A reviewer raised the interesting possibility that the predicted effects of substitutions we observe in Apocynaceae-feeding insects may in fact be ameliorated by substitutions elsewhere in the protein (e.g. compensatory substitutions).

In this second approach, we asked about the properties of ouabain-docking for eleven native proteins with ≥1 "large-effect" substitutions and eight with no such substitutions (Table S3). We approximated structures for each native protein using homology-based structural modeling and the PDB 3N23 crystal structure. Each native protein sequence was aligned to PDB 3N23 with MAFFT (44), MUSCLE (45) and 3D-coffee (46) using T-COFFEE EXPRESSO (47). These alignments and the PDB 3N23 structure were used as input for MODELLER v. 9.9 (48,49), which implements comparative protein structure modeling by satisfaction of spatial restraints (50,51). The quality of the models was assessed by DOPE (Discrete Optimized Protein Energy) score, GA341 score and the MODELLER objective function (molpdf, see MODELLER v.9.9 documentation). Models were structurally analyzed by visual inspection and prepared for figures in UCSF Chimera v. 1.5.3 (52,53). Molecular docking simulations of each of these predicted structures against ouabain were performed as described above. The RMSD from the PDB 3N23 docking and binding affinity of the top 20 dockings were assessed and used to determine the best ouabain docking position (as above).

We found a significant positive correlation between the RMSDs of the best native ouabain dockings and the total number of substitutions at sites implicated in ouabain-binding function (Spearman: ρ=0.587, P=0.007). The same trend can be seen for the affinities of the best dockings to the native proteins. Specifically, the affinity of the best docking is significantly negatively correlated with the total number of substitutions at sites implicated in ouabain-binding function (Spearman: ρ=0.561, P=0.0124). Additionally, the RMSDs of best native ouabain dockings are significantly larger for native proteins with one or more mutations of large effect (as determined in Table S2) compared to native proteins lacking such a substitution (Mann Whitney U-test: P=8e-4). These results link substitutions observed in Apocynaceae-feeding taxa (Fig. 1) with functional effects on ouabain binding in their native protein contexts.

Approach 3: As further validation of Approach 2, we investigated the ouabain-docking properties of the native protein modified to incorporate substitutions that revert observed substitutions at sites implicated in ouabain-binding function (Fig.1) to their consensus state (Table S4). Starting with the predicted structures of the native proteins for each taxa (see Approach 2 above), we introduced each substitution and compared the best docking of these "revertant" versions of the native proteins to the best docking for PDB 3N23 ATP1A1 (pig) and the original native protein (WT). We found that, revertant native proteins with at least one "large-effect" substitution (Table S2) bind significantly closer to the best pig docking relative to the true native proteins (11 closer: 0 farther, Sign test: P=0.001). A similar pattern is observed for the docking affinities of revertant native proteins. Specifically, revertant native proteins with at least one "large-effect" substitution bind with higher affinity to their best docking position relative to the true native proteins (10 higher: 1 lower, Sign test P=0.001). These results further link substitutions identified as having "large-effects" (Table S2) with functional effects on ouabain binding in their native protein contexts.

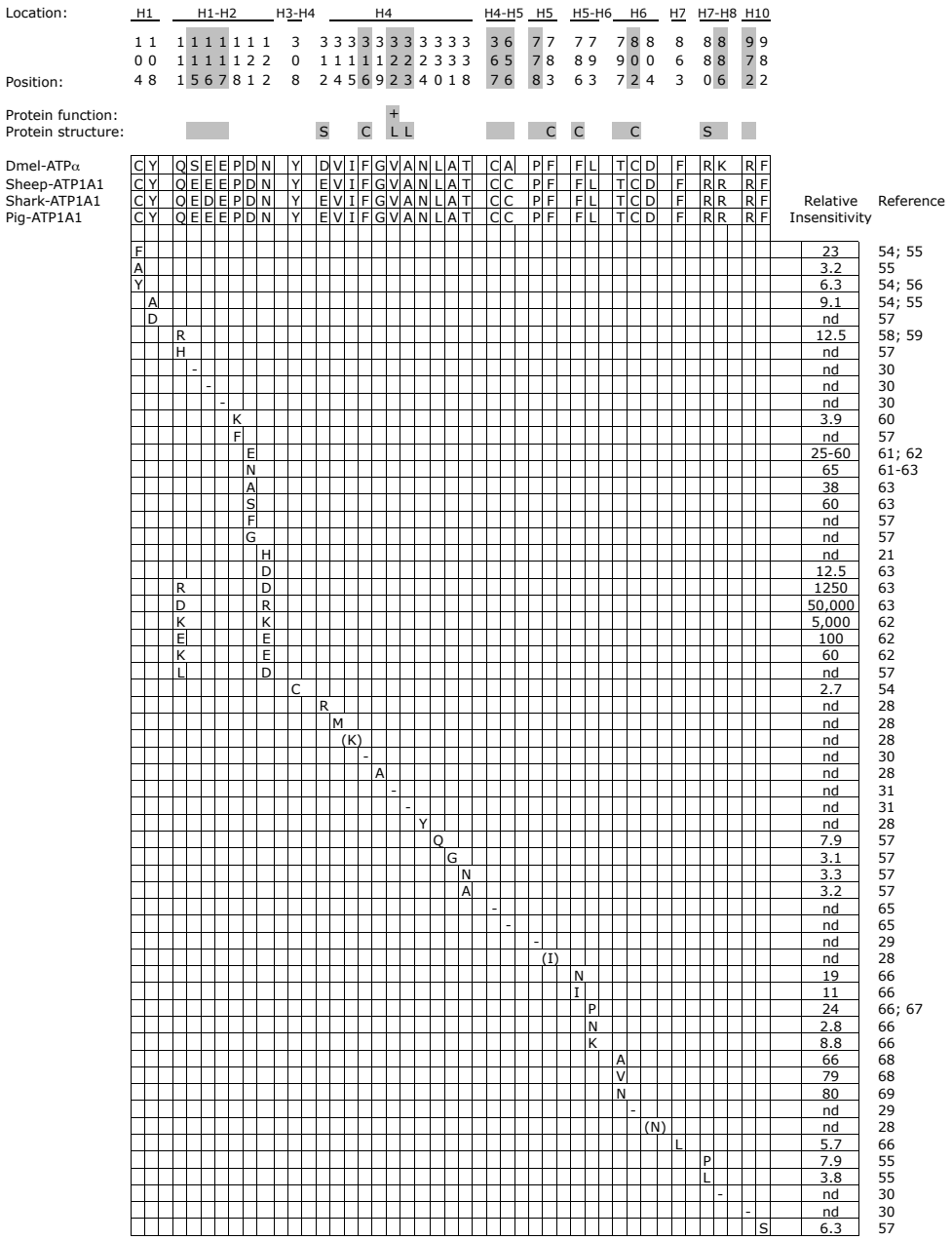
4. Copy-specific ATPα1 qPCR primers and their efficiencies.

Copy-specific cloning primers and qPCR primers (Table S9) were designed manually, based on assembled sequences for dogbane beetle (*C. auratus*, two copies), milkweed stem weevil (*R. lineaticollis*, two copies), large milkweed bug (*O. fasciatus*, three copies) and small milkweed bug (*L. kalmii*, three copies). Using the cloning primers, each of the ten distinct copies of ATPα1 was PCR-amplified individually from cDNA using PrimeSTAR HS DNA Polymerase (Takara). PCR products were gel-purified and cloned into a pCR™2.1-TOPO vector (TOPO Cloning kit, Invitrogen). Individual clones were sequenced to confirm the presence of the targeted ATPα1 copy. This set of ten plasmids was used as templates to test the specificity and efficiency of qPCR primers. Allele-specificity of qPCR primers was verified using reciprocal PCR trials on plasmids containing each copy. Efficiency of qPCR primers was determined using 10-fold serial dilutions. Cycling conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, and a 1 min annealing/extension step (60°C for *C. auratus*, 63°C for *R. lineaticollis*, 55°C for O. fasciatus, and 60°C for *L. kalmii*).

5. Tissue-specific expression of ATPa1 copies A, B and C in large and small milkweed bugs.

Motivated by the observation that putatively resistant and susceptible copies of ATPα1 are differentially expressed among tissues in the dogbane beetle and the milkweed stem weevil (Fig. 3), we set out to see if a similar pattern of expression is also found among duplicates found in the large and small milkweed bugs. In both beetle species, we found that the ratio of expression of ATPα1B (putatively resistant) relative to ATPα1A (putatively sensitive) follows the rank order (lowest to highest) head-muscle-gut. In large and small milkweed bugs, the situation is complicated by the fact that there are three copies of ATPα1 and all share a common substitution (N122H) that is known to confer some degree of insensitivity to ouabain. Despite this, we can rank the predicted sensitivity of these three duplicates in milkweed bugs based on the number of substitutions of known or predicted effects (Table S5). In particular, ATPα1A is the most derived at putatively functional sites with eight substitutions, five of which either have known effects on ouabain sensitivity or are predicted to affect ouabain-binding based on structural studies. In contrast, ATPα1C has two substitutions, only one of which with known effects on ouabain-sensitivity. The pattern of substitution at ATPα1B suggests it is intermediate between these extremes.

Despite the lack of a completely sensitive copy in milkweed bugs, we nonetheless predict that ATPα1A should have the lowest sensitivity among the three copies, whereas ATPα1C should have the highest sensitivity (Table S5). Thus, similar to the approach we took in beetles, we looked at tissue-specific expression of ATPα1A (least sensitive) vs ATPα1C (most sensitive) and ATPα1B (less sensitive) vs ATPα1C (most sensitive). We found a pattern of tissue-specific expression (Fig. S4) that closely mimics the pattern found in beetles (Fig. 3). In particular, the three copies show considerable differential expression among tissues with relative expression of the less sensitive copies to the most sensitive copy following the rank order (lowest to highest) head-muscle-gut. There is remarkable conservation of this tissue-specific differentiation pattern in the two species despite having diverged ~28 Mya (described above).



NOTE - Positions highlighted in gray are predicted to interact with ouabain solely based on protein structure data. A subset of these are labeled according to Ogawa et al. (29) as interacting with the steroid core (C), the sugar moiety (S), or the lactone ring (L). Site 322 is predicted to bind K+ (29). Relative Insensitivity denotes the fold-difference in sensitivity with the wild-type enzyme as measured in functional assays. nd = not determined.

Fig. S1

Summary of sites in ATPα involved in ouabain-binding based on site-directed mutagenesis studie and protein structure data.

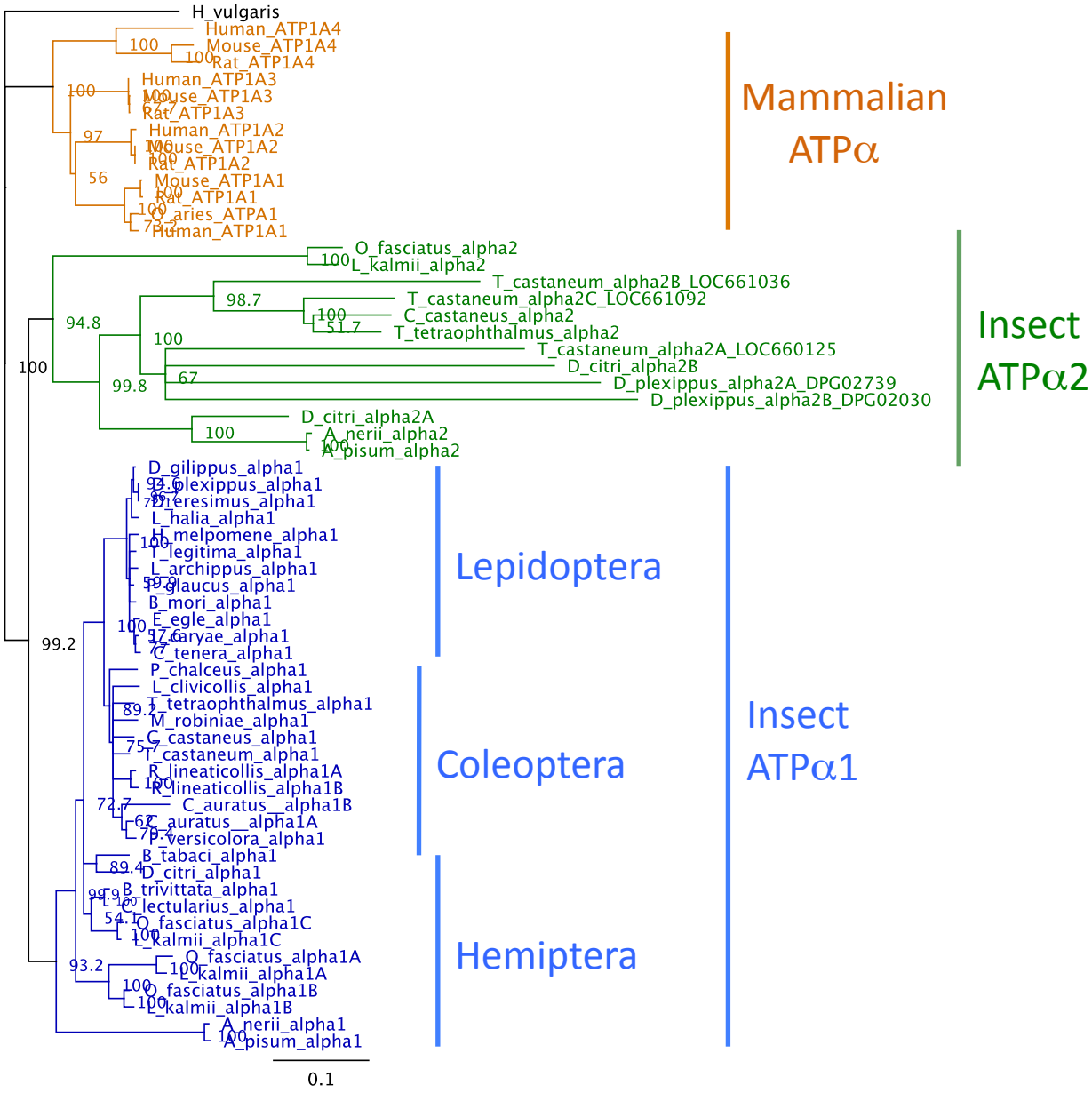
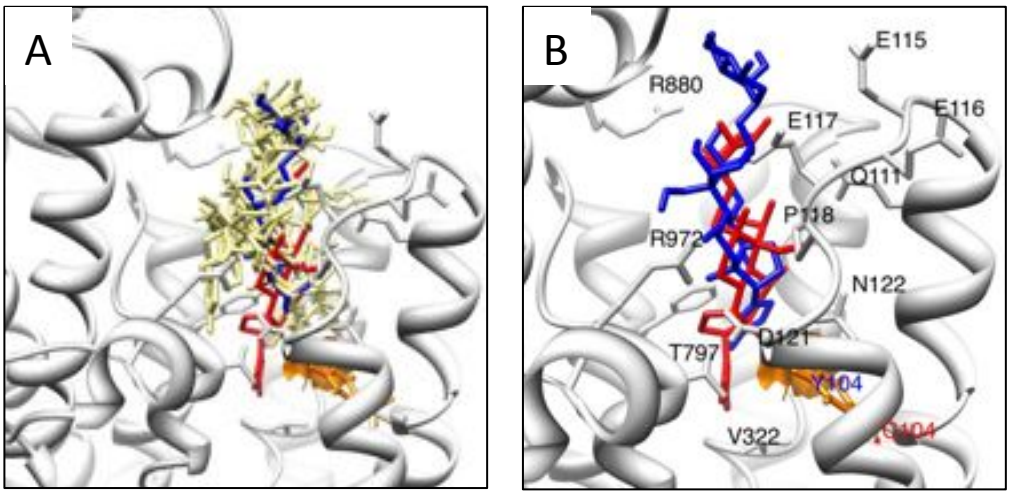


Fig. S2

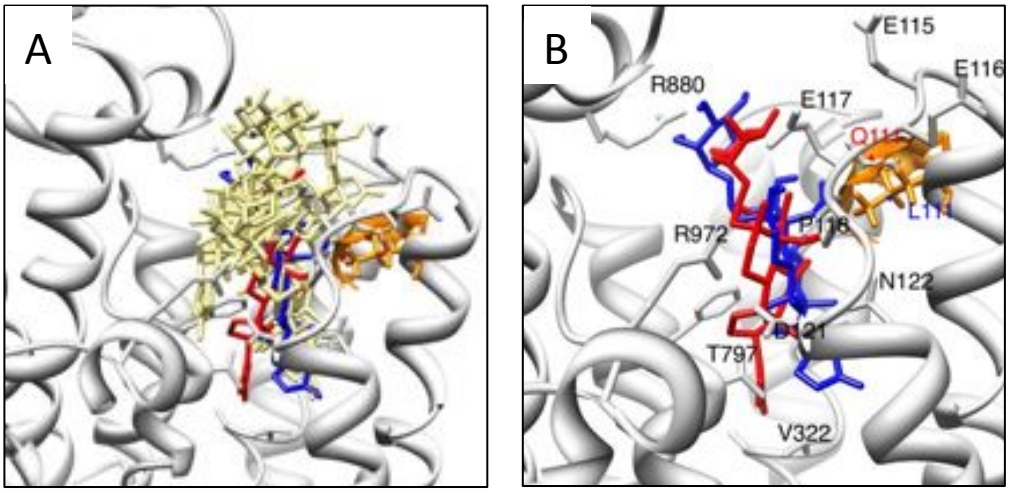
Phylogenetic relationships of ATPα orthologs and paralogs. Shown is a Neighbor-Joining tree for predicted protein sequences. Nodes are labeled with bootstrap percentages based on 1000 bootstrap replicates. The tree is rooted with ATPα of Cnidarian *Hydra vulgaris* (Genbank Acc # M75140). ATPα1 and ATPα2 of insects form a distinct clade from ATP1A1-4 of mammals. The deep divergence of ATPα1 and ATPα2 of insects and the presence of both duplicates in all three insect orders suggests this duplication arose before the diversification of insect orders (~300 Mya). For a list of taxa surveyed see Table S1.

Fig. S3

C104Y.



Q111V.



Q111L.

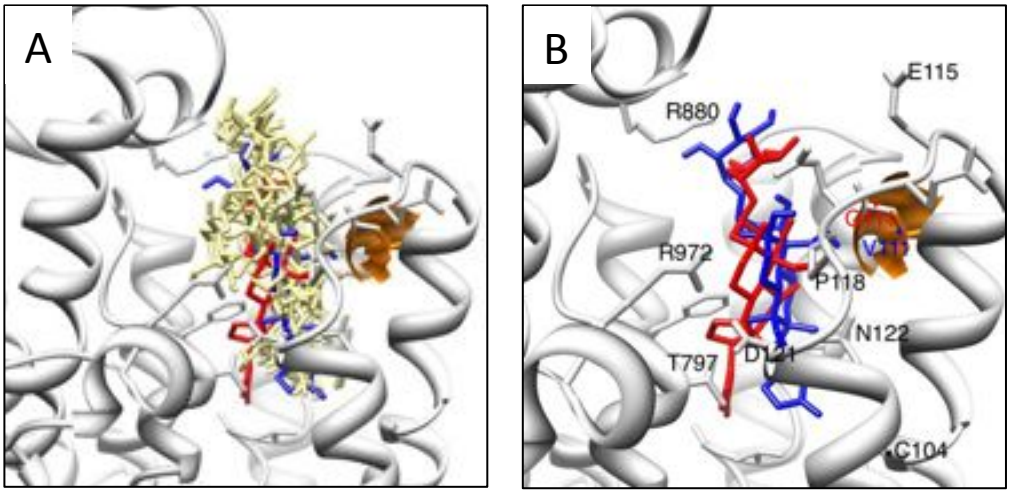
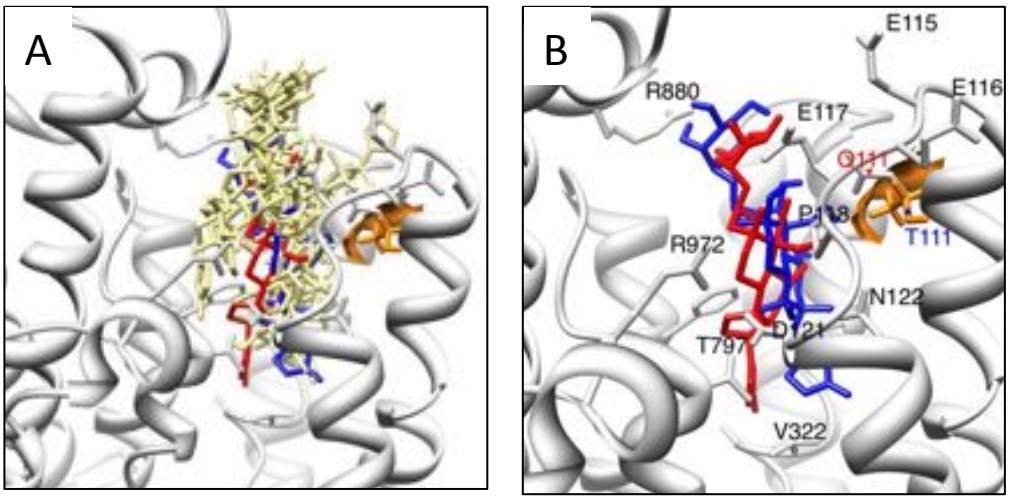
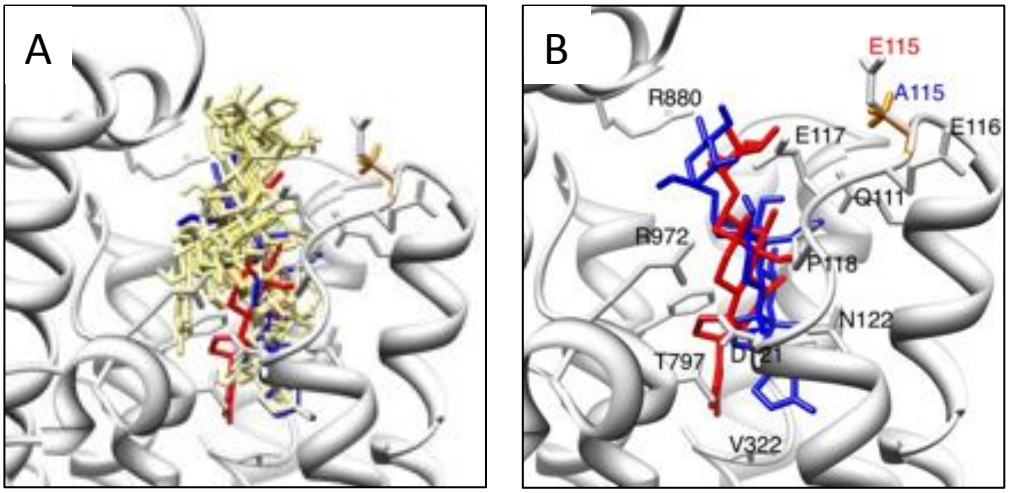


Fig. S3

Q111T.



E115A.



E115V.

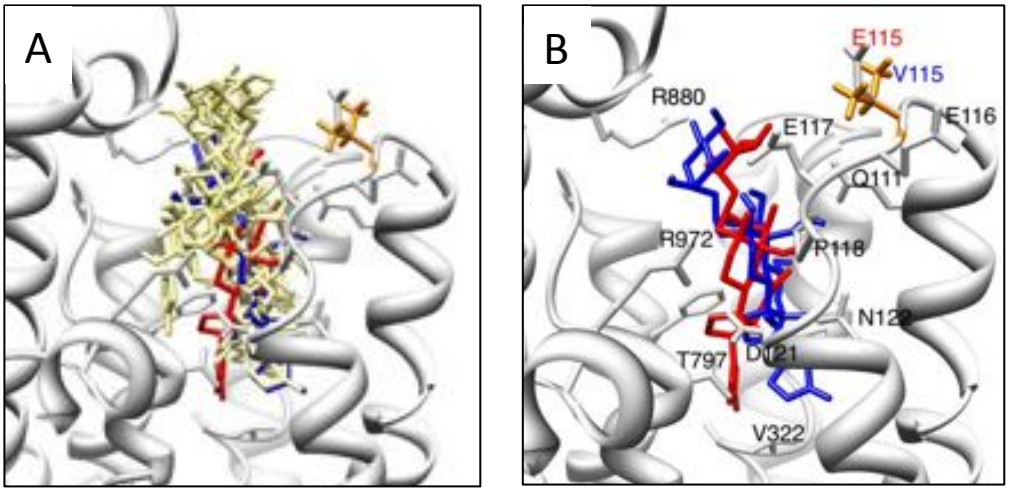
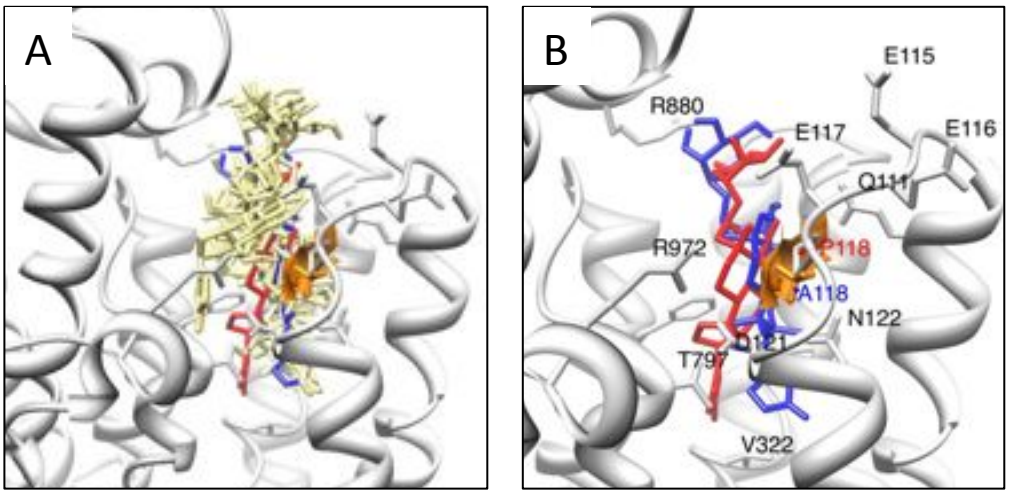
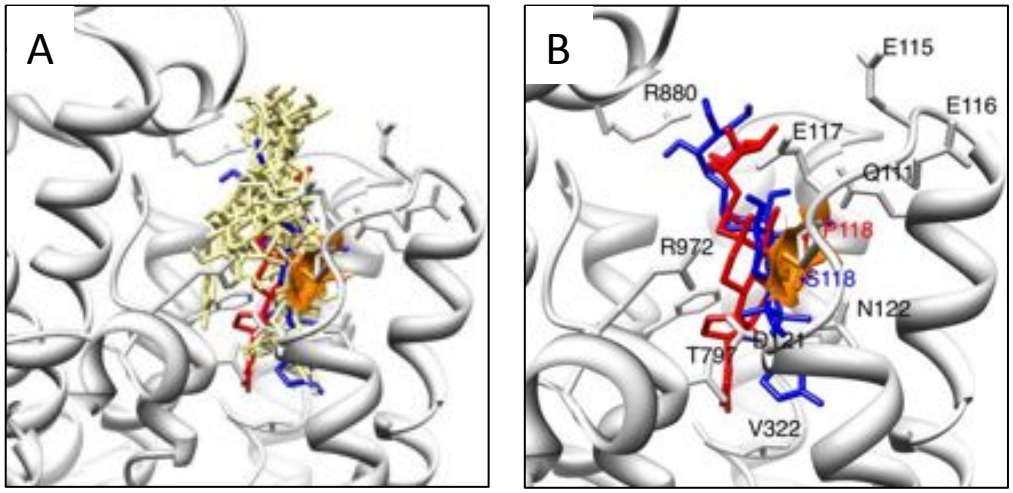


Fig. S3

P118A



P118S



D121N

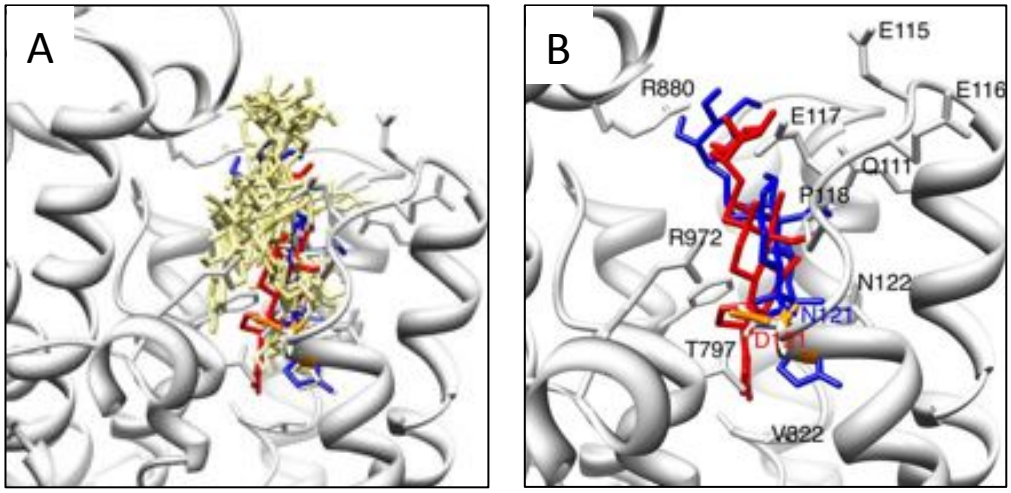
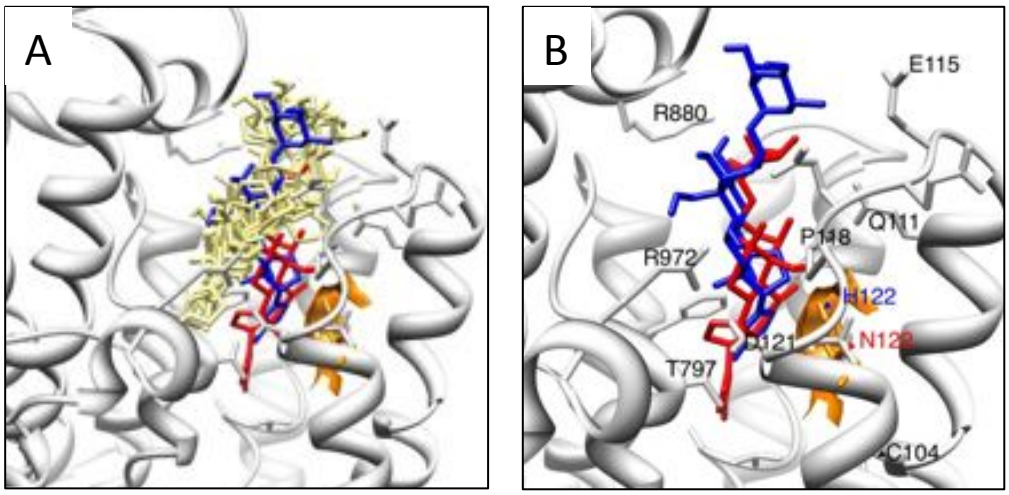
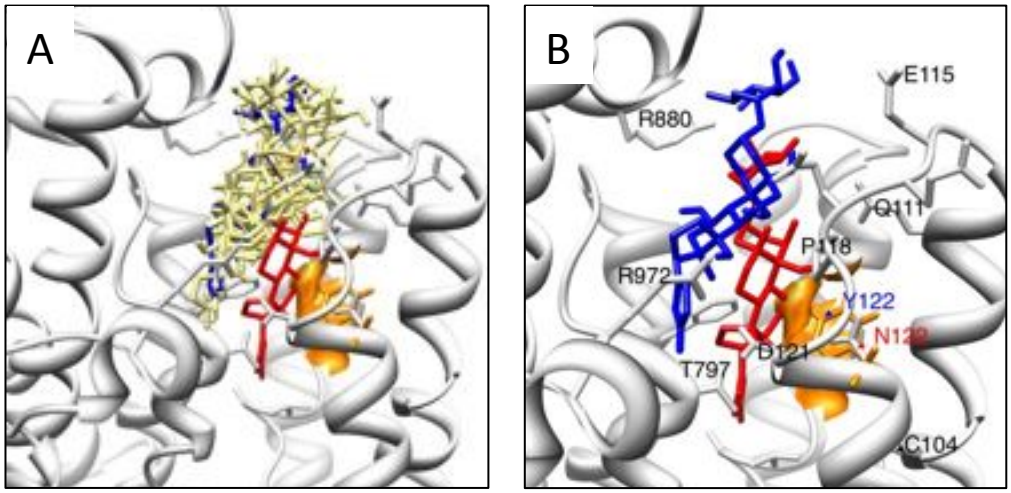


Fig. S3

N122H



N122Y



I315L

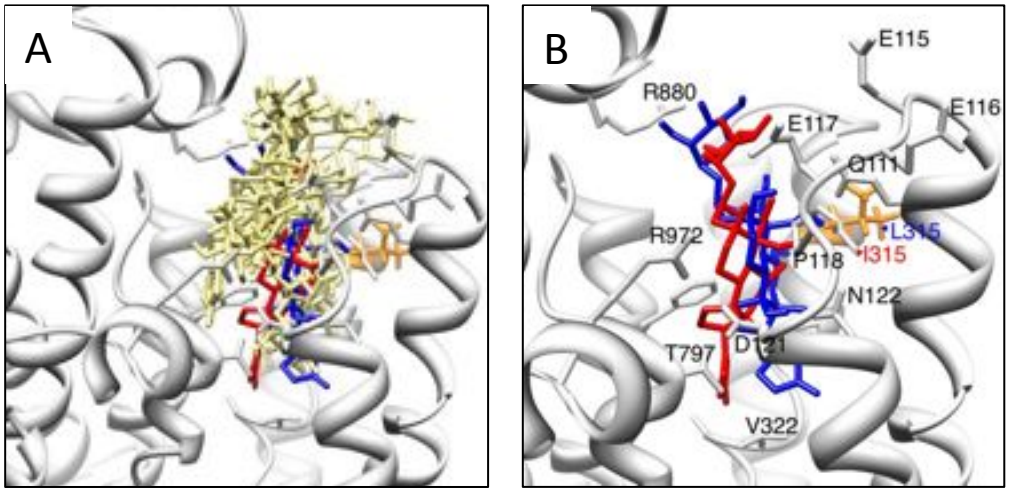
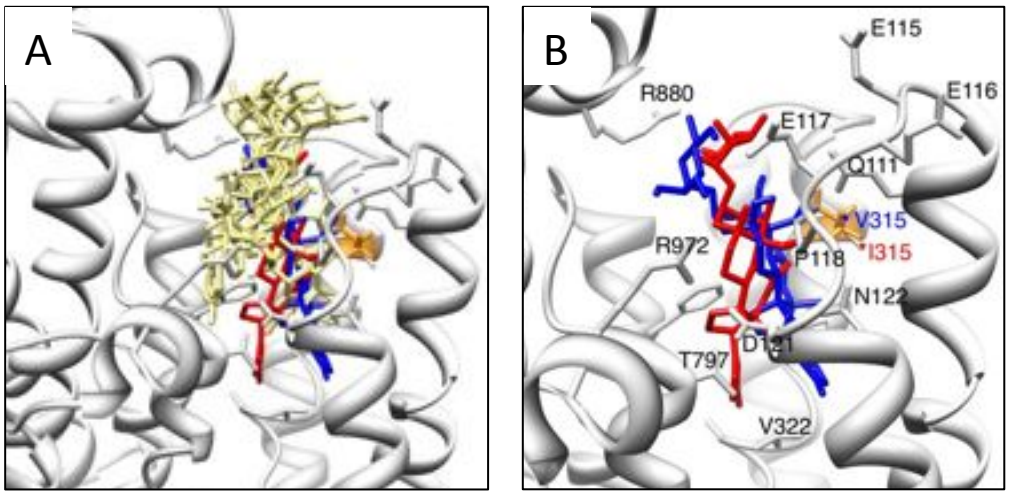
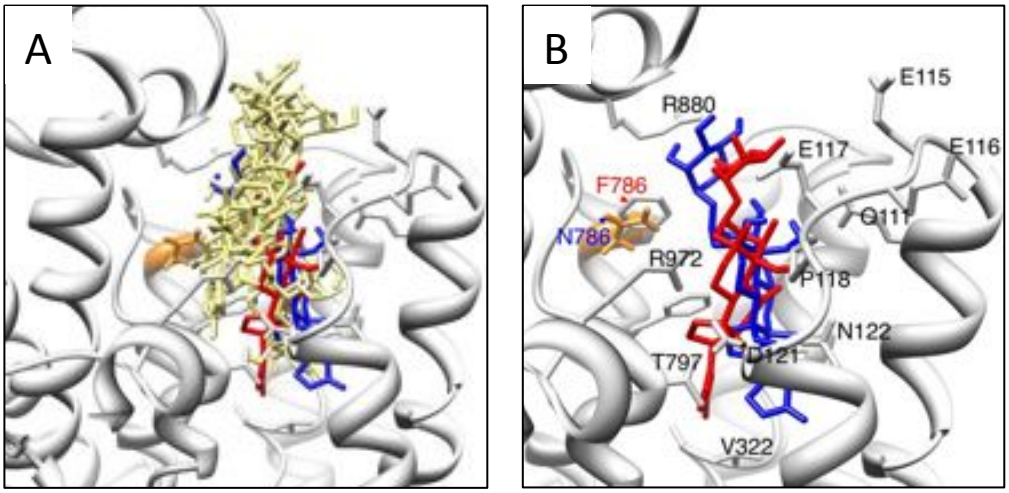


Fig. S3

I315V



F786N



T797A

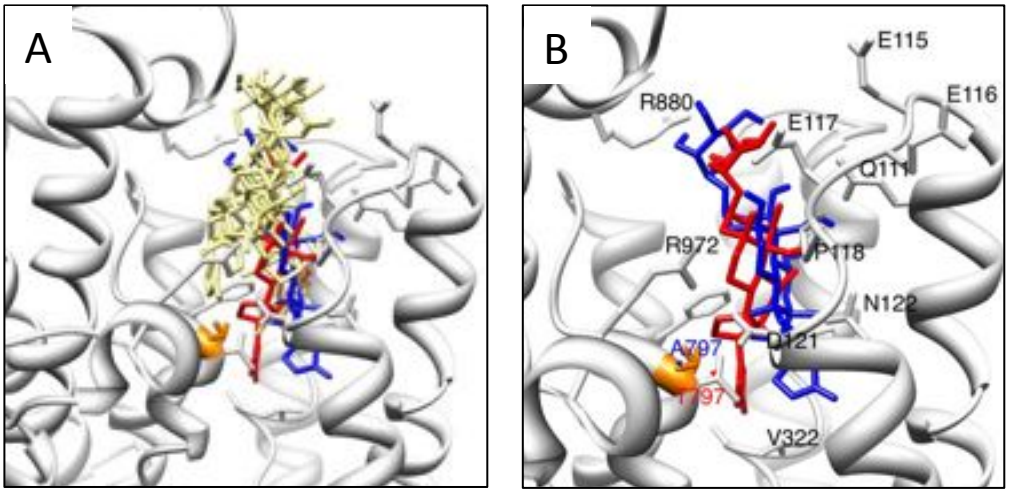
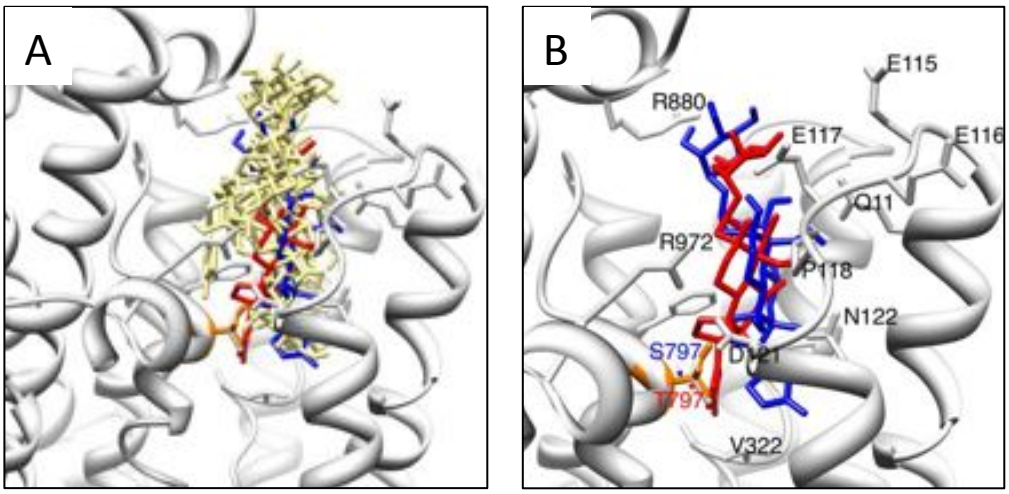
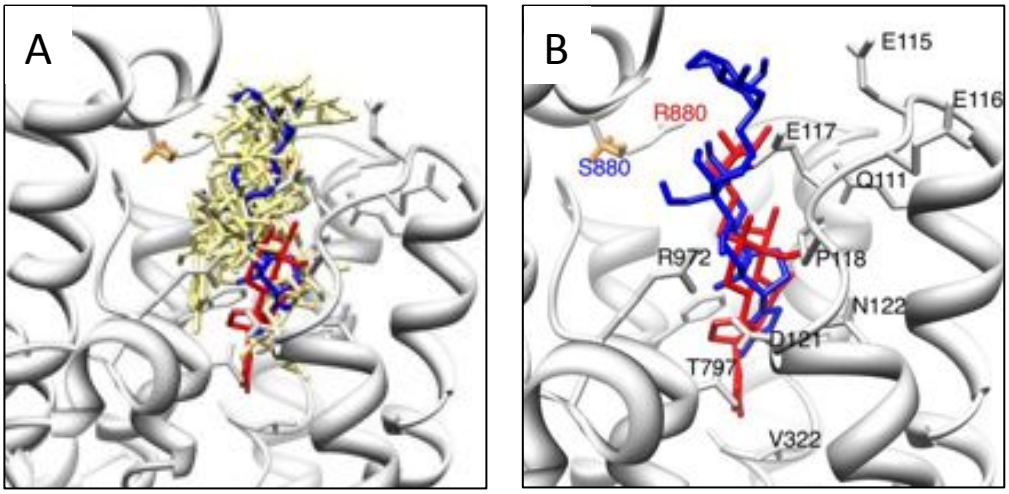


Fig. S3

T797S



R880S



R972Q

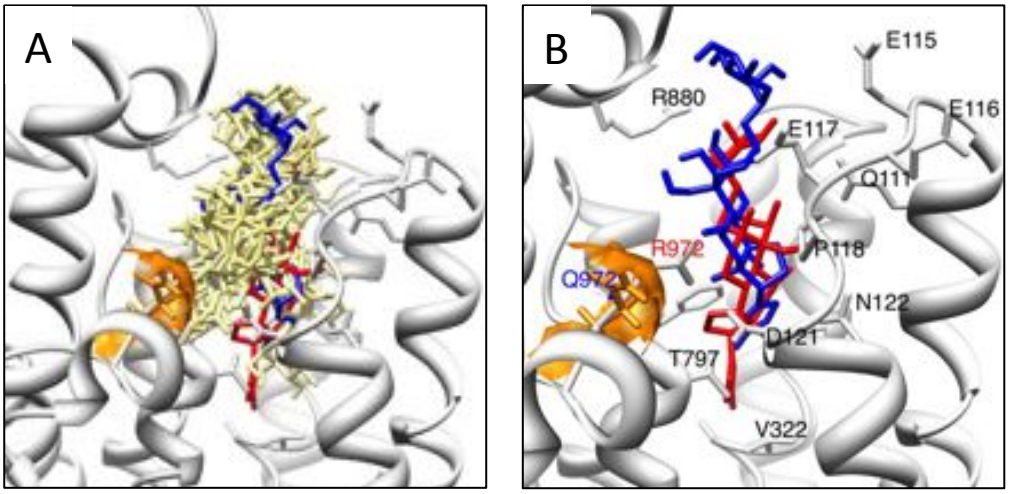


Fig. S3

WT

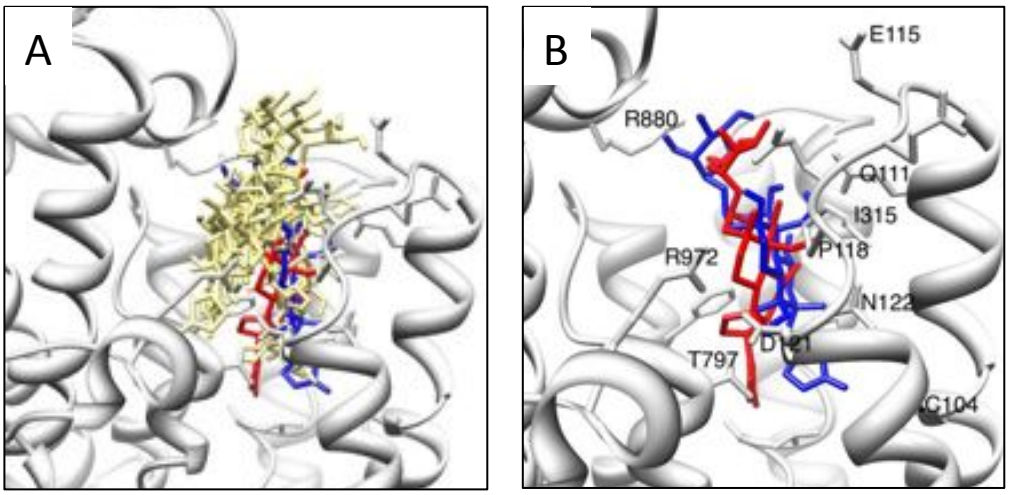


Fig. S3

Visual summary of molecular docking simulations for each of the substitutions in Table S2. (A) The best docking is shown in blue and is defined as closest to the coordinates of ouabain in the co-crystal structure (in red). The nine next best dockings are shown in yellow. For R972Q the number of dockings had to be increased to 20 visualize an ouabain docking with an affinity range close to the WT re- docking analysis. (B) A simplified version of the first panel that shows only the best docking onto the mutant protein (in blue) and the coordinates of ouabain in the co-crystal structure (in red). The atomic surfaces of the derived amino acid side chains are detailed in gold.

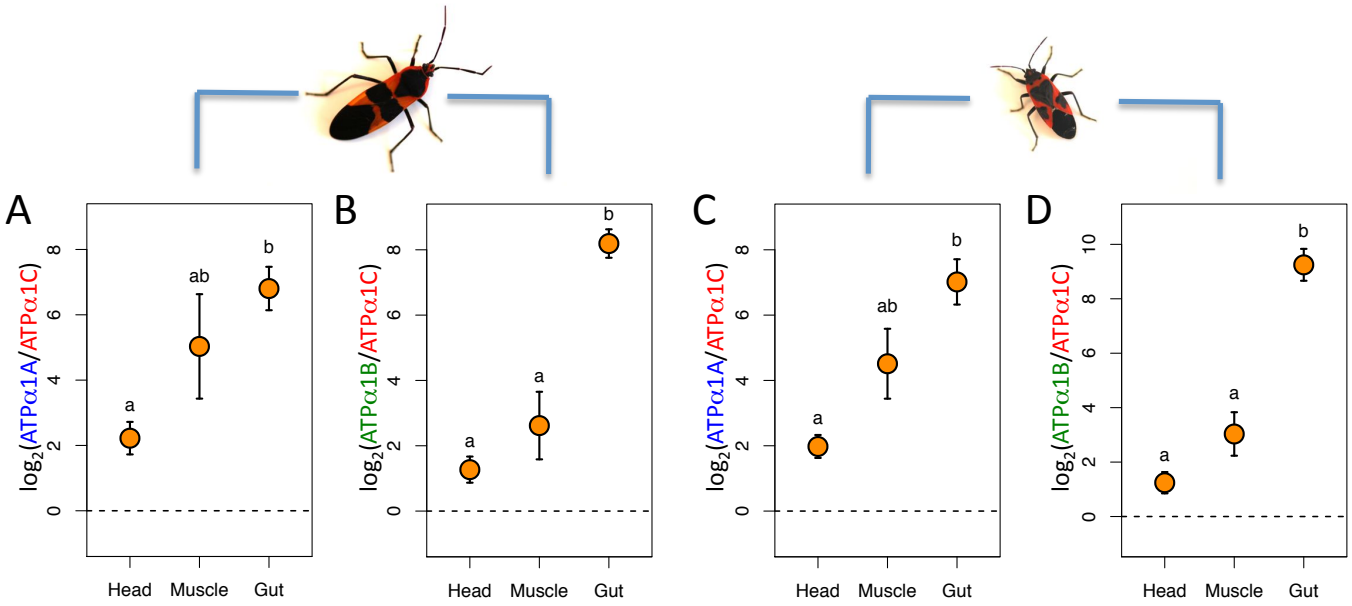


Fig. S4

Tissue-specific relative expression of ATPα1 duplicates in the large milkweed bug panels A and B) and the small milkweed bug (panels C and D). Shown are means with standard errors of biological replicates (four and five for the large and small milkweed bugs, respectively). In both cases, the expression of the putatively less-sensitive copy (ATPα1A or ATPα1B) is contrasted relative to the putatively most-sensitive copy (ATPα1C). Positive values indicate higher expression of the putatively resistant copy, ATPα1A or ATPα1B. Tukey-Kramer significance test categories are denoted with the letters a and b. In both cases, the relative expression of the putatively less-sensitive copy is significantly higher in the gut than in the head ($P=0.03$ panel A, $P=0.0001$ panel B, $P=0.0015$ panel C and $P<0.0001$ panel D), with muscle being intermediate between the two. The rank-order of relative expression level among tissues in both species mirrors a similar pattern observed in the dogbane beetle and the milkweed stem weevil (Fig. 3).

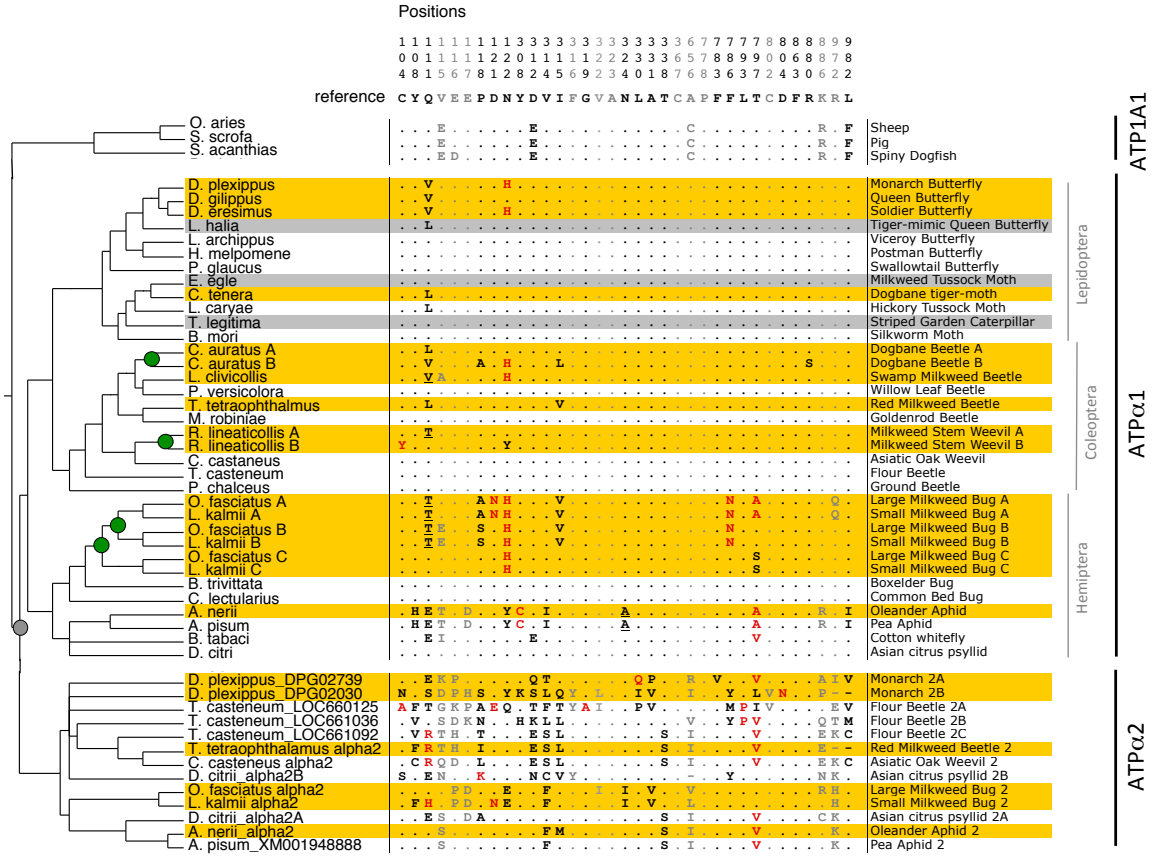


Fig. S5
Patterns of substitution at functional sites identified in mutagenesis and structural studies (Figure 1) extended to include ATPα2. Refer to Figure 1 legend for details on colors and shading. The ancient duplication that led to two ATPα lineages in insects is denoted by the grey circle.

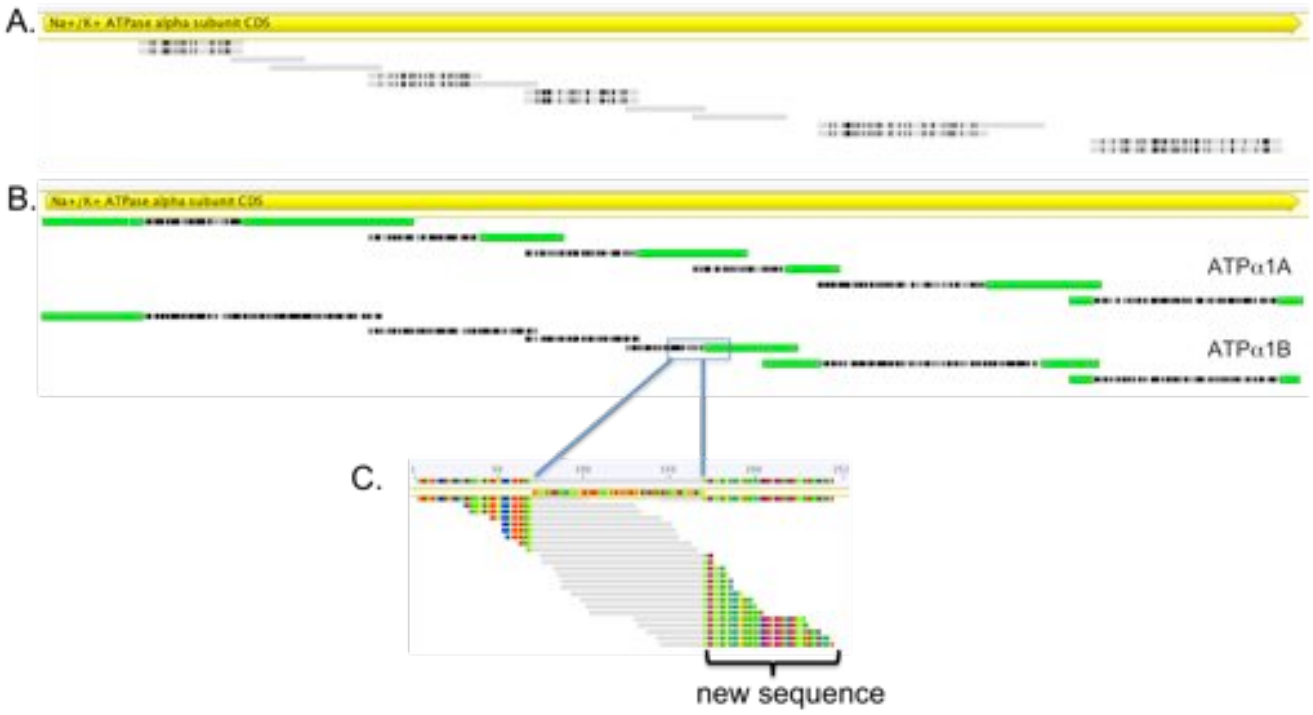


Fig. S6

Manual reconstruction of *C. auratus* ATPα1 copies. (A) Fourteen distinct Velvet contigs were identified with high sequence similarity to *D. melanogaster* ATPα1. (B) Two near full length ATPα1 orthologs were reconstructed by iterative manual extension of both ends of each contig until they overlapped. Manually extended regions are highlighted in green. (C) An example of manual contig-end extension: 100 bp sequences from each contig end were extracted and used as references (the gray portion between two solid blue lines) for assembly of all raw trimmed reads (only a subset of mapped reads are shown here). We accepted extensions into a gap that were supported by at least three reads.

Fig. S8

Establishing duplicates in *O. fasciatus*. Shown is the alignment of three copies of ATPα1 in *O. fasciatus*. Yellow bars indicate the position of CDS. Vertical black lines indicate nucleotide differences between copies. Red marks indicate polymorphic sites detected in each copy. All polymorphic sites were copy-specific.

Table S1
List of species and collection information.

Species	Common Name	Order	Family	Locality	Apocynaceae (Yes/No)
Danaus plexipus	Monarch Butterfly	Lepidoptera	Nymphalidae	Clinton Co, MI	Yes
Danaus gilippus	Queen Butterfly	Lepidoptera	Nymphalidae	Wacussa, FL	Yes
Danaus eresimus	Soldier Butterfly	Lepidoptera	Nymphalidae	El Azulillo, Mexico	Yes
Lycorea halia atergatis	Tiger-mimic Queen	Lepidoptera	Nymphalidae	El Azulillo, Mexico	Yes
Limenitis archippus	Viceroy	Lepidoptera	Nymphalidae	Clinton Co, MI	No
Papilio glaucus	Eastern Tiger Swallowtail	Lepidoptera	Papilionidae	Clinton Co, MI	No
Euchaetes egle	Milkweed Tussock Moth	Lepidoptera	Arctiidae	Princeton, NJ	Yes
Cycnia tenera	Dogbane tiger-moth	Lepidoptera	Arctiidae	Princeton, NJ	Yes
Lophocampa caryae	Hickory Tussock Moth	Lepidoptera	Arctiidae	New London Co, CT	No
Trichordestra legitima	Striped Garden Caterpillar	Lepidoptera	Noctuidae	Princeton, NJ	Yes[1]
Chrysochus auratus	Dogbane Beetle	Coleoptera	Chrysomelidae	Princeton, NJ	Yes
Labidomera clivicollis	Swamp Milkweed Leaf Beetle	Coleoptera	Chrysomelidae	Princeton, NJ	Yes
Plagiodera versicolora	Imported Willow Leaf Beetle	Coleoptera	Chrysomelidae	Princeton, NJ	No
Tetraopes tetraophthalmus	Red Milkweed Beetle	Coleoptera	Cerambycidae	Princeton, NJ	Yes
Megacyllene robiniae	Locust Borer Beetle	Coleoptera	Cerambycidae	Princeton, NJ	No
Rhyssomatus lineaticollis	Milkweed Stem Weevil	Coleoptera	Curculionidae	Princeton, NJ	Yes
Cyrtepistomus castaneus	Asiatic Oak Weevil	Coleoptera	Curculionidae	Kent Co, DE	No
Oncopeltus fasciatus	Large Milkweed Bug	Hemiptera	Lygaeidae	Princeton, NJ	Yes
Lygaeus kalmii	Small Milkweed Bug	Hemiptera	Lygaeidae	Princeton, NJ	Yes
Boisea trivittata	Eastern Boxelder Bug	Hemiptera	Rhopalidae	Princeton, NJ	No
Aphis nerii	Oleander Aphid	Hemiptera	Aphididae	Princeton, NJ	Yes

[1]Not typically associated with Apocynaceae, but occasionally found feeding on milkweeds (*70*) and collected *on Asclepias syriaca*

Table S2

Summary of effects of each substitution associated with use of Apocynaceae (see Fig. 1) on docking of ouabain onto the pig ortholog of ATPα (PDB 3N23A).

Substitution	Known effect of ouabain sensitivity?	RMSD[1] from co-crystal position in 3N23A (Å)	RMSD[1] from best WT docking position (Å)	Qualitative assessment of effect on ouabain binding.
C104Y	Yes	5.37	5.53	Large
Q111L	No	2.95	0.10	Subtle[2]
Q111V	No	2.98	0.14	Subtle[2]
Q111T	No	2.97	0.12	Subtle
E115A[3]	No	3.03	1.49	Subtle
E115V[3]	No	3.03	1.50	Subtle
P118A	No	2.94	0.09	Subtle
P118S	No	2.95	0.09	Subtle
D121N	Yes	2.92	0.03	Subtle[2]
N122H	Yes	5.49	5.64	Large
N122Y	No	6.40	6.94	Large
I315L	No	2.93	0.06	Subtle
I315V	No	3.48	1.88	Subtle
F786N	Yes	3.07	0.65	Subtle
T797A	Yes	2.93	0.04	Subtle[2]
T797S	No	2.92	0.04	Subtle
R880S	No	5.37	5.53	Large
R972Q	No	5.35	5.51	Large
None (WT)	-	2.93	-	-

NOTE – A substitution is deemed of large effect (highlighted in black) if the RMSD exceeds 4.6Å (the resolution of 3N23A).

[1] Root mean square distance. The best docking is defined as closest to the coordinates of ouabain in the co-crystal structure of (*31*).

[2] These substitutions disrupt a possible hydrogen bond between the protein and ouabain in the structure of Yatime et al. (*31*). D121N and T797A have known effects on ouabain-sensitivity (see Fig. S1).

[3] At this site, the state for pig 3N23A (E) differs from the consensus ancestral state for insects (V). Thus, these substitutions are modeled as E115A and E115V, instead of V115A and V115E, as observed in dogbane beetle and milkweed bugs, respectively (Fig. 1).

Table S3

Summary of dockings of ouabain onto predicted structures of native ATPα1 for eleven native proteins with ≥1 "large-effect" substitutions (blue) and eight with no such substitutions (black).

Native protein	Closest docking in top 20			
	RMSD(pig)	Affinity	#subs	#LEsubs
D. plexippus	4.23	-7.4	2	1
D. eresimus	**5.55**	-7.5	2	1
C. auratus B	**5.86**	-7.4	5	2
L. clivicolis	**6.72**	-8.1	3	1
R. lineaticollis B	**6.21**	-8.0	2	2
O. fasciatus A	1.57	-7.3	8	2
L. kalmii A	**5.03**	-7.5	8	2
O. fasciatus B	**4.97**	-7.1	6	1
L. kalmii B	4.43	-7.7	6	1
O. fasciatus C	**6.65**	-7.8	2	1
L. kalmii C	2.14	-8.0	2	1
D. gilippus	0.56	-8.0	1	0
L. halia	1.54	-7.5	1	0
E. egle	3.12	-8.2	0	0
C. tenera	1.15	-7.6	1	0
T. legitima	1.43	-7.9	0	0
C. auratus A	0.72	-7.9	1	0
T. tetrophthalmus	**4.92**	-7.6	2	0
R. lineaticollis A	1.00	-8.3	1	0

NOTE – RMSD(pig): Root-mean-square-distance in Angstroms to best docking of ouabain onto pig ATPA1 (PDB 3N23); #subs: Total number of substitutions observed at functionally important sites (Fig. 1); #LEsubs: The number of substitutions detected of "Large effect" on docking (Table S2). Bolded RMSD are >4.6 Angstroms, which is the resolution of the structure (PDB 3N23) on which these predicted structures are based.

Table S4

Summary of dockings of ouabain onto "revertant" versions of native ATPα1.

Native Protein	Substitutions incorporated	Closest docking in top 20 RMSD(pig)	ΔRMSD(pig)	ΔAffinity	RMSD(WT)
D. plexippus	V111Q + H122N	3.48	-0.75	-0.8	1.47
D. eresimus	V111Q + H122N	3.13	-2.42	-0.6	**4.89**
C. auratus B	V111Q + A118P + H122N + L315I + S880R	3.38	-2.48	-0.3	**7.55**
L. clivicolis	V111Q + A115V + H122N	0.66	-6.06	-0.4	**6.94**
R. lineaticollis B	Y104C + Y122N	0.53	-5.68	-0.1	**5.95**
O. fasciatus A	T111Q + A118P + N121D + H122N + V315I + 786F + S797T + Q972R	1.22	-0.35	-1.5	1.71
L. kalmii A	T111Q + A118P + N121D + H122N + V315I + 786F + S797T + Q972R	4.45	-0.58	-0.8	**4.62**
O. fasciatus B	T111Q + A115V+ S118P + H122N + V315I + N786F	**4.61**	-0.36	-0.4	**5.35**
L. kalmii B	T111Q + A115V+ S118P + H122N + V315I + N786F	0.92	-3.51	-0.5	4.39
O. fasciatus C	H122N + S797T	0.80	-5.85	-1.0	**6.75**
L. kalmii C	H122N + S797T	0.83	-1.31	0.1	2.63
D. gilippus	V111Q	0.81	0.25	0.2	0.45
L. halia	L111Q	4.34	2.80	-0.2	4.34
C. tenera	L111Q	1.12	-0.03	-0.4	0.07
C. auratus A	L111Q	3.85	3.13	0.1	1.68
T. tetrophthalmus	L111Q + L315I	2.87	-2.05	0.1	**5.86**
R. lineaticollis A	T111Q	**5.18**	4.18	-0.3	4.01

Note - Each native protein was modified to incorporate revertant substitutions corresponding to the substitutions documented in Fig. 1. RMSD(pig): Root-mean-square-distance in Angstroms to best docking of ouabain onto pig ATP1A1 (PDB 3N23); ΔRMSD(pig): change in RMSD with respect to the best docking for the native protein (Table S3). ΔAffinity: Change in affinity of best docking relative to the best docking for the native protein. RMSD(WT): Root-mean-square-distance in Angstroms to best docking of ouabain onto native protein (Table S3).

Table S5.

Functional substitutions observed in duplicates of ATPα1 of large and small milkweed bugs.

Copy	Substitutions of known effects on ouabain binding (Fig. S1)	Substitutions with predicted "Large effects" on ouabain binding (Table S2)	Substitutions at sites implicated in ouabain-binding (Fig. 1)	Expected Rank
ATPα1A	D121N, N122H, F876N, T797A	D121N, N122H, T797A, R972Q	Q111T, P118A, I315V	Lowest sensitivity
ATPα1B	N122H, F876N	N122H	Q111T, V115E, P118S, I315V	Intermediate sensitivity
ATPα1C	N122H	N122H	T797S	Highest sensitivity

Table S6
Taxa in this study with references for phylogenetic relationships.

	Scientific Name	Common Name	References for phylogeny construction
Lepidoptera	*Danaus plexippus*	Monarch Butterfly	71-74
	Danaus gilippus	Queen Butterfly	71-74
	Danaus eresimus	Soldier Butterfly	71-74
	Lycorea halia	Tiger-mimic Queen	71-74
	Limenitis archippus	Viceroy Butterfly	71-73
	Heliconius melpomene	Postman Butterfly	71-73
	Papilio glaucus	Eastern Tiger Swallowtail	71-73
	Euchaetes egle	Milkweed Tiger Moth	71; 73; 75; 76
	Cycnia tenera	Dogbane Tiger Moth	71; 73; 75; 76
	Lophocampa caryae	Hickory Tussock Moth	71; 73; 75; 76
	Trichordestra legitima	Striped Garden Caterpillar	71; 73; 76
	Bombyx mori	Silkworm Moth	71; 73; 76
Coleoptera	*Chrysochus auratus*	Dogbane Beetle	43; 73; 77
	Labidomera clivicollis	Swamp Milkweed Leaf Beetle	43; 73; 77
	Plagiodera versicolora	Imported Willow Leaf Beetle	43; 73; 77
	Tetraopes tetraophthalmus	Red Milkweed Beetle	73; 77
	Megacyllene robiniae	Locust Borer	73; 77
	Rhyssomatus lineaticollis	Milkweed Stem Weevil	73; 77
	Cyrtepistomus castaneus	Asiatic Oak Weevil	73; 77
	Tribolium casteneum	Red Flour Beetle	73; 77
	Pogonus chalceus	Saltmarsh Beetle	73; 77
Hemiptera	*Oncopeltus fasciatus*	Large Milkweed Bug	73; 78; 79
	Lygaeus kalmii	Small Milkweed Bug	73; 78; 79
	Boisea trivittata	Boxelder Bug	73; 78; 79
	Cimex lectularius	Common Bed Bug	73; 78; 79
	Aphis nerii	Oleander Aphid	73; 78; 80
	Acyrthosiphon pisum	Pea Aphid	73; 78; 80
	Bemisia tabaci	Whitefly	73; 78; 80
	Diaphorina citri	Asian Citrus Psyllid	73; 78; 80

Table S7
Details of mRNA-seq data collected.

Species	Common Name	Stage/Tissue	Single End		Paired End	
			# reads	length	# reads	length
Danaus plexipus	Monarch Butterfly	adult/thorax	28,266,296	101	-	-
Danaus gilippus	Queen Butterfly	adult/thorax	-	-	1,738,243	100-101
Danaus eresimus	Soldier Butterfly	adult/thorax	-	-	5,878,381	101
Lycorea halia atergatis	Tiger-mimic Queen	adult/thorax	16,067,453	98	1,450,610	100-101
Limenitis archippus	Viceroy	adult/thorax	3,933,071	100	-	-
Papilio glaucus	Eastern Tiger Swallowtail	adult/thorax	25,863,591	101	-	-
Euchaetes egle	Milkweed Tussock Moth	larvae/whole	5,905,128	95-100	6,988,976	101
Cycnia tenera	Dogbane tiger-moth	larvae/whole	-	-	5,629,493	101
Lophocampa caryae	Hickory Tussock Moth	larvae/whole	4,958,026	100	-	-
Trichordestra legitima	Striped Garden Caterpillar	larvae/whole	-	-	4,737,299	101
Chrysochus auratus	Dogbane Beetle	adult/head+thorax	3,284,909	95	3,888,265	100-101
Labidomera clivicollis	Swamp Milkweed Leaf Beetle	adult/head+thorax	6,935,044	95-100	-	-
Plagiodera versicolora	Imported Willow Leaf Beetle	adult/whole	1,907,103	100	-	-
Tetraopes tetraophthalmus	Red Milkweed Beetle	adult/head+thorax	8,273,718	95-100	-	-
Megacyllene robiniae	Locust Borer Beetle	adult/head+thorax	4,245,152	100	-	-
Rhyssomatus lineaticollis	Milkweed Stem Weevil	adult/whole	3,417,256	95	4,297,782	100-101
Cyrtepistomus castaneus	Asiatic Oak Weevil	adult/whole	-	-	12,392,116	101
Oncopeltus fasciatus	Large Milkweed Bug	adult/whole	7,888,794	95-100	27,109,674	101
Lygaeus kalmii	Small Milkweed Bug	adult/whole	9,476,280	95-100	21,224,112	101
Boisea trivittata	Eastern Boxelder Bug	adult/head+thorax	4,858,096	100	-	-
Aphis nerii	Oleander Aphid	adults(clone)/whole	-	-	14,098,396	101

Table S8
List of reference genome sequences and accession numbers used in this study.

Species	Common Name	Order	Family	Accession number	Apocynaceae (Yes/No)
Ovis aries	Sheep	Artiodactyla	Bovidae	NM_001009360	No
Sus scrofa	Pig	Artiodactyla	Suidae	PDB: 3N23_A	No
Squalus acanthias	Spiny dogfish	Squaliformes	Squalidae	PDB: 3A3Y_A	No
Bombyx mori	Domestic silk worm	Lepidoptera	Bombycidae	Silkbase: BGIBMGA005058-TA	No
Danaus plexippus	Monarch butterfly	Lepidoptera	Nymphalidae	DPGLEAN09831, AGBW01004098.1	No
Heliconius melpomene	Postman butterfly	Lepidoptera	Nymphalidae	SRA: SRR349678	No
Drosophila melanogaster	Fruit fly	Diptera	Drosophilidae	AF044974, NZ_AABU00000000.1	No
Tribolium castaneum	Red flour beetle	Coleoptera	Tenebrionidae	XM_969867, AAJJ00000000.1	No
Pogonus chalceus	Ground beetle	Coleoptera	Carabidae	SRA: SRR424342	No
Bemisia tabaci	Sweet potato whitefly	Hemiptera	Aleyrodidae	SRA: SRX018661	No
Diaphorina citri	Asian citrus psyllid	Hemiptera	Psyllidae	SRA: SRR063791	No
Acyrthosiphon pisum	Pea aphid	Hemiptera	Aphididae	XM_001948888, ABLF00000000.2	No
Cimex lectularius	Common bedbug	Hemiptera	Cimicidae	SRA: SRR349675	No

Table S9
Summary of copy-specific ATPα1 qPCR primers and their qPCR efficiencies.

	Forward primer	Reverse primer	Efficiency
C. auratus			
ATPα1A	AGATATAATGAAGCGAAGACCT	GTATTCCAATGTTTTCCTATCACGA	95%
ATPα1B	ATAATGAAGAGGCCCCCA	GAATTCCAACATTTTTCTGTCATTG	97%
R. lineaticollis			
ATPα1A	GTAAAAACCTATTCGGCGGATTT	AACACAAYACCCAAGAACAAATA	97%
ATPα1B	GTAAAAATCTATTYGGYGGTTTC	AACACAACWCCCAAGAACAGATT	100%
O. fasciatus			
ATPα1A	GACCCAAAGCAAAAGAGCGAAG	CAAGAGATGGGCCTAATGCATTAACA	102%
ATPα1B	GAGATGGGCCTAACGCTTTAACC	GCCAACCCACAACAATAAAGCAAAA	95%
ATPα1C	GCGATGGACCTAATGCACTTACT	TCCAACCCATAGTAGAAGTGCGAAT	100%
L. kalmii			
ATPα1A	AACATGGTTCCACAGTTTGCTTG	CTCCTAAAACAATTGCCTCTGCA	101%
ATPα1B	ATGGTGCCGCAGTTTGCTGT	CTCCTAAAACTATTTCCTCTGCG	103%
ATPα1C	GGTGCCCCAGTTTGCCAC	CCCCAGTACAATGTCTTCTGCT	103%

Table S10

Proportion of total branch length sampled for taxa "using" and "not using" Apocynaceae host plants.

Group	Using (dS*S)	Using (%)	Not using (dS*S)	Not using (%)
Lepidoptera	1159.1	31.4	2529.3	68.6
Coleoptera	3475.1	39.7	5270.0	60.3
Hemiptera	3155.7	24.5	9742.7	75.5
All	7789.9	30.8	17542.0	69.2
All (excluding Sternorrhyncha)	7700.0	48.1	8298.6	51.9

NOTE – dS is the per site rate of substitution at synonymous sites and S is the number of synonymous sites. dS*S corresponds to the estimated number of synonymous substitutions along a lineage.

Table S11

Of lineages associated with Apocynaceae host plants, the proportion of total branch length sampled for lineages with "duplicated" and "not duplicated" versions of ATPα1.

Group	Duplicated (dS*S)	Duplicated (%)	Not duplicated (dS*S)	Not duplicated (%)
Lepidoptera	0	0	1159.0	100
Coleoptera	1327.9	38.2	2147.2	61.8
Hemiptera (minus Sternorrhyncha)	3065.8	100	0	0
All	4393.7	57.1	3306.2	42.9

Supplementary References

36. M. F. Polz, C. M. Cavanaugh, *Appl. Environ. Microbiol.* **64**, 3724 (1998).
37. J. G. Gibbons *et al.*, *Molecular biology and evolution* **26**, 2731 (2009).
38. Z. Yang, *Molecular biology and evolution* **24**, 1586 (2007).
39. O. Trott, A. J. Olson, *J. Comput. Chem.* **31**, 455 (2010).
40. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).
41. J. M. Wang, W. Wang, P. A. Kollman, D. A. Case, *J. Mol. Graph. Model* 25, 247 (2006).
42. D. R. Zerbino, E. Birney, *Genome Res* 18, 821 (2008).
43. J. Gomez-Zurita, T. Hunt, F. Kopliku, A. P. Vogler, *PloS one* 2, e360 (2007).
44. H. Kim, S. Lee, Y. Jang, *PloS one* 6, e24749 (2011).
45. K. Katoh, K.-i. Kuma, H. Toh, T. Miyata, *Nucleic Acids Res.* 33, 511 (2005).
46. R. C. Edgar, *BMC Bioinformatics* 5, 1 (2004).
47. C. Notredame, D. G. Higgins, J. Heringa, *J. Mol. Biol.* 302, 205 (2000).
48. F. Armougom *et al.*, *Nucleic Acids Res.* 34, W604 (2006).
49. N. Eswar *et al.*, *Current Protocols in Protein Science*. (John Wiley & Sons, Inc., 2007), pp. 2.9.1-2.9.31.
50. M. A. Marti-Renom *et al.*, *Annu. Rev. Biophys. Biomolec. Struct.* 29, 291 (2000).
51. A. Fiser, R. K. G. Do, A. Sali, *Protein Sci.* 9, 1753 (2000).
52. A. Sali, T. L. Blundell, *J. Mol. Biol.* 234, 779 (1993).
53. E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, T. E. Ferrin, *BMC Bioinformatics* **7**, 339 (2006).
54. C. M. Canessa, J. D. Horisberger, D. Louvard, B. C. Rossier, *Embo J* **11**, 1681 (1992).
55. P. J. Schultheis, E. T. Wallick, J. B. Lingrel, *The Journal of biological chemistry* **268**, 22686 (1993).
56. G. R. Askew, J. B. Lingrel, *Journal of Biological Chemistry* **269**, 24120 (1994).
57. M. L. Croyle, A. L. Woo, J. B. Lingrel, *European Journal of Biochemistry* **248**, 488 (1997).
58. E. M. Price, J. B. Lingrel, *Biochemistry* **27**, 8400 (1988).
59. V. Canfield, J. R. Emanuel, N. Spickofsky, R. Levenson, R. F. Margolskee, *Mol Cell Biol* **10**, 1367 (1990).
60. P. J. Schultheis, J. B. Lingrel, *Biochemistry* **32**, 544 (1993).
61. E. M. Price, D. A. Rice, J. B. Lingrel, *The Journal of biological chemistry* **264**, 21902 (1989).
62. E. M. Price, D. A. Rice, J. B. Lingrel, *The Journal of biological chemistry* **265**, 6638 (1990).
63. J. B. Lingrel, J. Orlowski, M. M. Shull, E. M. Price, in *Progress in Nucleic Acid Research and Molecular Biology,* E. C. Waldo, M. Klvle, Eds. (Academic Press, 1990), vol. Volume 38, pp. 37-89.

64. C. M. Canessa, J. D. Horisberger, B. C. Rossier, *The Journal of biological chemistry* **268**, 17722 (1993).
65. T. L. Kirley, M. Peng, *The Journal of biological chemistry* **266**, 19953 (1991).
66. M. Palasis, T. A. Kuntzweiler, J. M. Arguello, J. B. Lingrel, *The Journal of biological chemistry* **271**, 14176 (1996).
67. E. L. Burns, R. A. Nicholas, E. M. Price, *Journal of Biological Chemistry* **271**, 15879 (1996).
68. J. N. Feng, J. B. Lingrel, *Biochemistry* **33**, 4218 (1994).
69. E. L. Burns, E. M. Price, *Journal of Biological Chemistry* **268**, 25632 (1993).
70. H. B. Weiss, E. L. Dickerson, *Journal of the New York Entomological Society* **29**, 123 (1921).
71. N. P. Kristensen, M. J. Scoble, O. Karsholt, *Zootaxa*, 67, **669** (2008).
72. N. Wahlberg *et al.*, *Proceedings. Biological sciences / The Royal Society* **276**, 4295 (2009).
73. B. M. Wiegmann *et al.*, *BMC biology* **7**, 34 (2009).
74. A. V. Z. Brower, N. Wahlberg, J. R. Ogawa, M. Boppre, R. I. Vane-Wright, *Syst Biodivers* **8**, 75 (2010).
75. J. M. Ratcliffe, M. L. Nydam, *Nature* **455**, 96 (2008).
76. R. Zahiri *et al.*, *Zool Scr* **40**, 158 (2011).
77. T. Hunt *et al.*, *Science* **318**, 1913 (2007).
78. C. D. von Dohlen, N. A. Moran, *Journal of molecular evolution* **41**, 211 (1995).
79. H. M. Li *et al.*, *Molecular phylogenetics and evolution* **37**, 313 (2005).
80. M. L. Thao, L. Baumann, P. Baumann, *Bmc Evolutionary Biology* **4**, (2004).